



**Miguel Monsanto
Pinheiro**

Sistema de rastreio de doenças metabólicas



**Miguel Monsanto
Pinheiro**

Sistema de rastreio de doenças metabólicas

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Electrónica e Telecomunicações, realizada sob a orientação do Doutor José Luís Oliveira, Professor Associado do Departamento de Engenharia Electrónica e Telecomunicações da Universidade de Aveiro.

o júri

presidente

Prof. Dr. Armando José Formoso de Pinho
professor associado do Departamento de Electrónica e Telecomunicações
Universidade de Aveiro

Prof. Dr. Orlando Manuel de Oliveira Belo
professor associado do Departamento de Informática da Escola de Engenharia
Universidade do Minho

Prof. Dr. Manuel António da Silva Santos
professor auxiliar do Departamento de Biologia da
Universidade de Aveiro

Prof. Dr. José Luís Guimarães Oliveira
professor associado do Departamento de Electrónica e Telecomunicações
Universidade de Aveiro

agradecimentos

A todos aqueles que de alguma forma contribuíram para a elaboração desta dissertação o meu muito obrigado.

Professor Doutor José Luís Oliveira;
Instituto de Genética Médica – Dr. Hugo Rocha;
Teresa Monteiro;
Gaspar Dias;
Miguel Mussa;
Caetano Ferreira;
Rui Teixeira.

Aos meus pais. A eles dedico este trabalho.

palavras-chave

bioinformática, rastreio de neo-natais, doenças metabólicas.

resumo

O objectivo principal deste trabalho é desenhar e implementar um sistema de suporte à decisão, apoiado num conjunto de ferramentas matemáticas e estatísticas, que permita rastrear doenças metabólicas em recém-nascidos. O sistema terá que reduzir ao máximo o número recém-nascidos com diagnósticos de falsos positivos e negativos. Para tal, o sistema apoia-se na análise dos valores obtidos em recém-nascidos saudáveis/validados para obter, através do cálculo dos percentis, os valores máximos e mínimos dos limites necessários ao diagnóstico.

A concepção da interface está baseada num paradigma familiar que permite ao utilizador um manejo amigável e de fácil utilização, permitindo um rápido rastreio.

O sistema de informação proporciona também um controlo diário sobre as amostras submetidas, possibilitando a detecção de valores anormais originados por algum mau manuseamento dos espectrómetros ou uma anómala preparação das amostras para análise.

keywords

bioinformatics, newborn screening, metabolic diseases.

abstract

The main goal of this work is the development of an information system supported by a set of mathematical and statistical tools necessary to the screening of metabolic diseases in new born offspring. The system will have to minimize the number of newborns with fake positive and negative diagnosis. To achieve this goal, the system is based upon the analysis of the obtained values of healthy newborns in order to get, through the calculation of the percentis, the maximum and minimum values of the necessary limits to the diagnosis. The layout of the operator interface is constructed upon familiar paradigms allowing for increased user friendliness. The information system also permits a daily control of the submitted samples, allowing the detection of abnormal values originated by some improper usage of the mass spectrometers or an unsuitable preparation of the samples to be analyzed.

Índice

1	Introdução	1
1.1	Enquadramento	2
1.2	Objectivos.....	2
1.3	Estrutura da dissertação	3
2	Espectrometria de massa	5
2.1	Enquadramento histórico	5
2.2	Espectrómetro de massa	7
2.2.1	Introdução da amostra.....	8
2.2.2	Métodos de ionização da amostra.....	8
2.2.3	Análise e separação de iões.....	11
2.2.4	Detecção de iões.....	13
2.3	Espectrómetros de massa em tandem	14
2.3.1	Analisar com o espectrómetro de massa em tandem.....	14
2.4	Aplicações	16
2.5	Sumário	18
3	Conceitos biológicos	21
3.1	Células.....	21
3.2	Biologia molecular	23
3.3	Doenças Metabólicas	27
3.4	Sumário	30
4	Rastreio de doenças em recém-nascidos.....	33
4.1	Recursos (tecnologia, ferramentas, sistemas de informação.....)	33
4.2	Caso Português	37
4.3	Lacunas/problemas.....	39
4.4	Sumário	41
5	NeoScreen - ferramenta de auxílio ao diagnóstico de doenças metabólicas em recém-nascidos.	43
5.1	Requisitos.....	43
5.1.1	Descrição dos actores	44

5.1.2	Descrição dos casos de utilização.....	45
5.2	Base de dados	48
5.3	Implementação.....	54
5.4	NeoScreen – Interface de utilizador.....	61
5.5	Sumário	69
6	Resultados.....	71
7	Conclusões e trabalho futuro	77
8	Anexos	79
9	Referências	93

Lista das Figuras

Figura 1 – Aparelho desenvolvido por Thomson	6
Figura 2 - Divisão do espectrómetro de massa em três componentes	7
Figura 3 - Método de ionização por electrospray.	9
Figura 4 - Formação de iões pelo método de ionização por electrospray.	10
Figura 5 - Formação de iões através do método de ionização MALDI.....	11
Figura 6 – Analisador quadrupolo.....	12
Figura 7 - Espectrómetro de massa TOF acoplado a um MALDI.	12
Figura 8 - Multiplicador de electrões.....	13
Figura 9 - Espectrómetro de massa MS/MS composto por dois quadrupolos	14
Figura 10 - Modo de análise através de <i>Product ion Scanning</i>	15
Figura 11 - Modo de análise através de <i>Precursor ion scanning</i>	15
Figura 12 - Modo de análise através de <i>Constant neutral loss</i>	16
Figura 13 - Modo de análise através de <i>Multiple Reaction Monitoring</i>	16
Figura 14 - Espectros de massa obtidos para cada composto isolado	18
Figura 15 – Visão geral de uma célula eucariótica.....	22
Figura 16 - Exemplo de um segmento de ADN.	23
Figura 17 - Tabela de correspondência de codões para aminoácidos.....	24
Figura 18 – Todo o processo de transformação de um gene em proteína.....	25
Figura 19 - Processo de tradução do ARN mensageiro.	26
Figura 20 – Processo de rastreio dos recém-nascidos.....	35
Figura 21 – Concentrações obtidas na análise realizada a 40 indivíduos.....	40
Figura 22 – Duas formas de rastreio possível.....	41
Figura 23 – Diagrama de casos de utilização para o sistema NeoScreen.	44
Figura 24 – Conceito de limite que se aplica aos marcadores	46
Figura 25 – Diagrama conceptual da base de dados	49
Figura 26 - Ferramenta de assistência ao motor de base de dados.	53
Figura 27 – Grupo de classes impostas pela arquitectura MFC.....	55
Figura 28 - Arquitectura implementada no desenvolvimento do sistema.....	54
Figura 29 – Agregação de classes à classe CMainFrame.....	56
Figura 30 - Agregação de classes à classe CMetabolicView	57
Figura 31 – Estrutura de dados implementada através de classes	58
Figura 32 – Classes agregadas à classe CMetabolicDoc.....	60
Figura 33 – Janela principal da aplicação NeoScreen.	61
Figura 34 – Cartas de controlo	62
Figura 35 – Processo de leitura e rastreio dos indivíduos	63
Figura 36 - Definição dos marcadores.	64
Figura 37 - Definição de espectrómetros de massa.....	65
Figura 38 - Definição de controlos CDC.....	65

Figura 39 - Definição doenças.....	66
Figura 40 - Definição de regras.....	67
Figura 41 - Controlo dos marcadores através das médias diárias.....	68
Figura 42 – Estatísticas obtidas através dos histogramas.	68
Figura 43 – Multiposto através do NeoScreen.....	69
Figura 44 – Análise de uma carta de controlo.	74
Figura 45 – Evolução dos recém-nascidos identificados com patologias.....	75

Lista das Tabelas

Tabela 1 - Lista dos vários grupos de proteínas existentes.....	27
Tabela 2 - doenças rastreáveis em Portugal recorrendo ao (MS/MS).....	29
Tabela 3 – Número de casos detectados em 2002, 2003 e 2004	30
Tabela 4 – Descrição das entidades.....	50
Tabela 5 – Descrição das entidades fracas	50
Tabela 6 - Descrição dos relacionamentos com atributos	51
Tabela 7 – Descrição dos relacionamentos sem atributos.....	51
Tabela 8 – Valores calculados através dos percentis do IGM	72
Tabela 9 - Valores calculados através dos percentis do HCUS.....	73
Tabela 10 – Aminoácidos e suas abreviaturas	81
Tabela 11 – Valores previstos por cada dia de processamento.....	89
Tabela 12 – Contém as entidades envolvidas na base de dados.....	90
Tabela 13 – Entidades com crescimentos em função dos dias.	91

Lista de acrónimos

3-HMG	Acidúria 3-Hidroxi-3-MetilGlutárica.
ADN	Ácido desoxirribonucleico.
ARNm	Ácido ribonucleico mensageiro.
ARNt	Ácido ribonucleico de transferência.
ATP	Trifosfato de adenosina.
CDC	Centers for Disease Control and Prevention.
Da	Dalton unidade de massa normalmente usada para expressar a massa de um átomo ou de uma molécula; é aproximadamente a massa de um único próton ou neutrão e o seu equivalente é aproximadamente de 1.66×10^{-27} Kg.
ESI	Ionização por electrospray.
eV	Electrão Volt energia que um electrão ou próton ganha quando passa através de uma diferença de potencial de um volt.
FAB	Bombardeamento com átomos rápidos.
GC	Cromatografia gasosa.
GC/MS	Cromatografia gasosa / espectrómetro de massa.
HC	Hipotiroidismo congénito.
HCUS	Hospital Clínico Universitário de Santiago.
HPLC	Cromatografia líquida de alta desempenho.
MSUD	Leucínose.
LC	Cromatografia líquida.
LCAD	Deficiência nos ácidos gordos de cadeia longa.
IGM	Instituto de Genética Médica.
MALDI	Desadsorção com laser assistida por matriz.
MCADD/MCAD	Deficiência nos ácidos gordos de cadeia média
MS	Espectrómetro de Massa.
MS/MS	Espectrómetros de massa em tandem.
PKU	Fenilcetonúria.
SCAD	Deficiência nos ácidos gordos de cadeia curta.
TOF	Tempo de voo.

1 Introdução

O presente trabalho enquadra-se no âmbito das actividades do grupo de Bioinformática da Universidade de Aveiro.

Este grupo tem por objectivo desenvolver aplicações nas áreas da biologia molecular e da biomedicina de forma a solucionar alguns dos problemas que os investigadores se deparam na análise de grandes quantidades de informação.

Actualmente, o grupo tem em curso o desenvolvimento de quatro projectos distintos, sendo eles:

- NeoScreen – sistema de informação que dá suporte à decisão no rastreio neo-natal de doenças metabólicas, e que dá enquadramento à presente tese;
- DiseaseCard – sistema de informação que possibilita a partilha de conhecimento de doenças genéticas raras, através de um sítio na web, onde vários especialistas poderão organizar variada informação dispersa por múltiplas bases de dados públicas e privadas, estando posteriormente acessível ao utilizador comum;
- Mind – desenvolvimento e implementação de um sistema responsável pelo desenho e tratamento de dados gerados por microarrays. Estes chips são lâminas de vidro ou nylon, onde podem ser comparados de uma só vez milhares de genes ou fragmentos de genes, e são responsáveis por gerar grandes quantidades de informação;
- Anaconda – sistema de informação que permite analisar genomas, possibilitando a detecção de regras gerais que envolvem a tradução de um determinado gene em proteína. Para tal, foram desenvolvidas metodologias matemáticas e algoritmos para análise de grandes volumes de informação genética e implementados interfaces amigáveis para visualização da informação.

O autor esteve envolvido ao longo dos dois últimos anos no sistema Anaconda e na aplicação NeoScreen.

1.1 Enquadramento

O rastreio dos erros inatos do metabolismo em recém-nascidos, amplamente conhecido como o teste do pezinho, adiciona inúmeras vantagens à população em geral. Contribui para uma melhoria significativa do nível de vida, através do rastreio de doenças que só iriam manifestar-se numa fase posterior. O rastreio permite a detecção e um tratamento de algumas doenças de forma correcta e eficaz.

Nos meados dos anos noventa foi possível aumentar o número de doenças com possibilidade de detecção através do rastreio neo-natal. Para tal, muito contribuíram os progressos conseguidos no melhoramento dos espectrómetros de massa, utilizados para identificar e quantificar vários tipos de compostos presentes no sangue, recolhido pelos centros da saúde e hospitais aos primeiros dias de vida de um recém-nascido. Estes compostos, estando envolvidos em vários processos metabólicos permitem detectar, através da concentração presente no sangue, possíveis disfunções que se irão manifestar posteriormente ou, pode resultar mesmo no óbito do recém-nascido. Mediante os níveis de concentrações de determinados compostos presentes no sangue é possível determinar se determinado indivíduo é possuidor de determinada doença relacionada com o composto ou compostos em causa.

Em Portugal, em 2004, o rastreio cingia-se a duas doenças. No entanto, com a compra de dois espectrómetros de massa passou a ser possível aumentar o número de doenças rastreáveis.

A introdução desta tecnologia veio trazer também novos desafios, pois as análises realizadas através de espectrómetros produzem grandes quantidades de dados. É necessário possuir um sistema de informação capaz de gerir todos os dados, armazená-los e criar algoritmos que permitem detectar os recém-nascidos que potencialmente poderão padecer de alguma doença.

1.2 Objectivos

A dissertação surge integrada num projecto realizado entre Instituto Engenharia Electrónica e Telemática de Aveiro (IEETA) e o Instituto de Genética Médica Abel Salazar do Porto (IGM), responsável pelo rastreio de todos recém-nascidos em Portugal.

As necessidades do IGM devem-se ao facto de este possuir dois espectrómetros de massa necessários ao rastreio dos neo-natais, mas não existir capacidade de rastreio devido ao sistema produzir uma elevada quantidade de dados. A necessidade de desenvolver um sistema de informação, apoiado em

ferramentas matemáticas e estatísticas, que permitisse o rastreio e armazenamento dos dados gerados pelos espectrómetros de massa aquando da análise de sangue dos recém-nascidos, está na origem da presente tese.

A finalidade principal do sistema de informação, que efectue rastreios, é a de gerar o mais baixo número possível de falsos positivos e falsos negativos. Tem que se garantir que as crianças com doenças são detectadas e as que crianças sãs não sejam diagnosticadas como doentes. Esta questão é a mais importante a ter em conta num sistema de rastreio.

Assim, uma das questões mais importantes do projecto é aplicar ferramentas computacionais que retirem informação necessária à detecção de novas doenças, ainda não diagnosticadas, e diminuir o número de falsos positivos e negativos.

1.3 Estrutura da dissertação

Nos três primeiros capítulos são abordados os fundamentos associados a um programa de rastreio de doenças metabólicas. O teor destes capítulos afasta-se um pouco do domínio da tecnologia onde esta dissertação se insere. No entanto, o seu conhecimento é fundamental para o desenvolvimento da aplicação apresentada.

O capítulo dois introduz os conceitos principais associados à espectroscopia de massa, tecnologia que neste momento é fundamental para a obtenção dos dados necessários para efectuar rastreio. Serão apresentados vários tipos de espectrómetros, bem como os elementos necessários ao seu funcionamento.

O capítulo três centra-se nos conceitos biológicos que estão na origem das doenças metabólicas, que serão diagnosticadas pelo Programa Nacional de Rastreio Neo-Natal. Começa com a descrição das células, passando pelos conceitos genéticos e finalizando com uma descrição das doenças que serão rastreadas.

No capítulo quatro descrevem-se os métodos envolvidos no rastreio neo-natal, desde que a amostra sanguínea é obtida, no centro de saúde, até à obtenção do resultado final da análise. Descrevem-se, também, os cuidados a ter com as amostras, bem como, as recomendações dadas pelas entidades credenciadoras do rastreio neo-natal.

O capítulo cinco contém os passos envolvidos na criação da aplicação NeoScreen, passando pelos requisitos, diagramas de casos, base de dados e finalizando com uma descrição da aplicação. É feita, também, uma pequena abordagem às classes e suas dependências.

No capítulo seis apresentam-se os resultados obtidos com a utilização da aplicação, por parte dos laboratórios envolvidos, bem como uma breve discussão dos resultados.

O capítulo sete conclui esta dissertação resumindo os resultados e os ganhos obtidos nos laboratórios que utilizam a aplicação. São também apresentadas descritas algumas ideias e direcções possíveis para investigação futura.

2 Espectrometria de massa

O espectrómetro de massa é um instrumento usado para medir massas e concentrações relativas de átomos e moléculas. É utilizado para identificar moléculas desconhecidas, quantificar moléculas conhecidas e elucidar as propriedades químicas e estruturais das mesmas.

Cada molécula tem uma massa específica demasiado pequena para ser pesada pelos métodos convencionais. Para tal, recorre-se ao espectrómetro de massa para pesar as moléculas electronicamente. Além de identificar as moléculas pela sua massa, o espectrómetro quantifica a concentração relativa de uma determinada molécula presente na solução.

O sistema apresenta um elevado nível de sensibilidade, capaz de detectar um determinado tipo de molécula, existente numa solução, com uma concentração de 10^{-15} mol/ μ L, sendo possível obter um espectro de massa num curto espaço de tempo [1].

A espectrometria de massa é usada nos mais variados campos de pesquisa [2]:

- Biotecnologia - análise de proteínas, oligo-nucleótidos;
- Ambiente – qualidade de águas, contaminação de alimentos;
- Geologia – análise de crude;
- Saúde – rastreio de neo-natais, testes de drogas e estupefacientes.

2.1 Enquadramento histórico

Os fundamentos científicos que permitiram a utilização da espectrometria de massa tiveram origem em 1886, quando Goldstein detectou os iões positivos, tendo Wien, em 1898, analisado o mesmo através de deflexões magnéticas [3]. Mas em 1897, um físico inglês, Joseph John Thomson (1856-1940), demonstrou a existência do electrão, uma partícula sub-atómica parte constituinte de toda a matéria e responsável pelo transporte de correntes eléctricas [4]. Posteriormente, este cientista mostrou a existência da relação entre o electrão e a electricidade. Muitos outros fizeram descobertas no campo da electricidade, mas foi Thomson que provou ser a corrente eléctrica constituída por fluxos de pequenas partículas (electrões), transportando cada electrão uma carga eléctrica (Figura 1). Em 1906, Thomson recebeu o prémio Nobel da física pelo trabalho *“Investigations on passage of electricity through gases”*. Demonstrou que um

electrão é uma partícula com uma determinada carga e mediu o rácio entre massa/carga da mesma (m/z) [4].

Posteriormente, em 1913, editou um livro, no qual representava os espectros de massa do O_2 , N_2 , CO e CO_2 [5], estando lançado o caminho para a espectroscopia de massa.

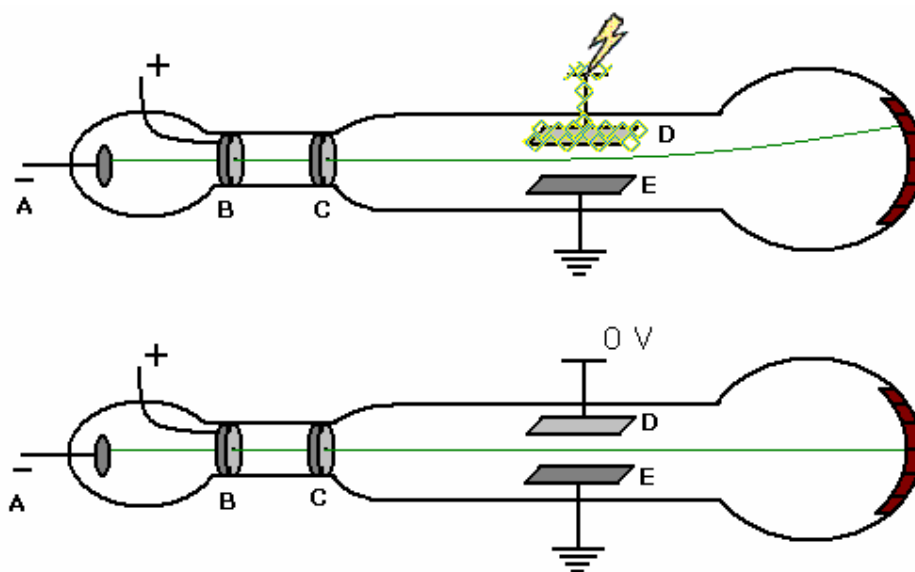


Figura 1 – Aparelho desenvolvido por Thomson para efectuar as suas experiências. Os electrões têm origem no cátodo aquecido, passando por uma fresta no ânodo B e passando, novamente, pela fresta da placa C. Quando uma diferença de tensão é aplicada às placas D e E provoca uma mudança de direcção na trajectória dos electrões. Por fim, os electrões colidem com um detector para medir a deflexão. Este aparelho é conhecido por tubo de raios catódicos.

Para perceber melhor os desafios da espectroscopia de massa é apropriado explicar alguns aspectos da experiência de Thomson.

Uma das primeiras condições para o sucesso da mesma era o vácuo. Vivemos num ambiente gasoso compostos por várias moléculas. Nas medições, Thomson assumiu que, somente o campo eléctrico afectava a trajectória dos feixes de electrões não devendo estes ser afectados na sua trajectória pelo choque com moléculas/átomos presentes no meio. Houve necessidade de remover o máximo de atmosfera possível para reduzir esse efeito ao máximo. Presentemente, o vácuo é uma condição obrigatória nos espectrómetros de massa.

Outro aspecto a ter conta era a criação de um feixe de electrões com uma velocidade conhecida. Para Thomson, criar uma nuvem de electrões era relativamente fácil, necessitando, somente, aquecer um fio a altas temperaturas. Para criar o feixe com uma dada direcção e com uma velocidade conhecida, Thomson aplicou um campo eléctrico de modo a acelerar os electrões gerados

pelo ânodo aquecido. Quanto maior é o campo, maior é a velocidade, podendo mesmo ser calculada com precisão. Este princípio é, ainda hoje, usado para criar feixes de iões nos espectrómetros de massa. Por último, o detector era constituído por uma grelha de fosforescência que emite um fotão por embate de electrão.

É este tipo de medidas realizadas pelos espectrómetros de massa, nos dias de hoje. Não medem os rácios massa/carga dos electrões, mas sim os rácios de átomos e moléculas com carga positiva/negativa, mais conhecidos por iões. Os fundamentos gerais da física aplicada aos espectrómetros de massa são os mesmos desenvolvidos por Thomson há 100 anos atrás.

2.2 Espectrómetro de massa

O espectrómetro de massa pode ser dividido em três partes fundamentais, nomeadamente fonte de iões, analisador e detector.

A amostra para investigação é introduzida na fonte de iões do espectrómetro. Uma vez introduzida na fonte de iões, as moléculas são ionizadas e aceleradas por campos eléctricos para dentro do analisador. Dentro do analisador, os iões são desviados da sua trajectória consoante a sua massa/carga (m/z). Posteriormente, os iões são detectados e o sinal gerado é enviado para um sistema computacional, onde ocorre a construção do espectro de massa da solução introduzida (Figura 2).

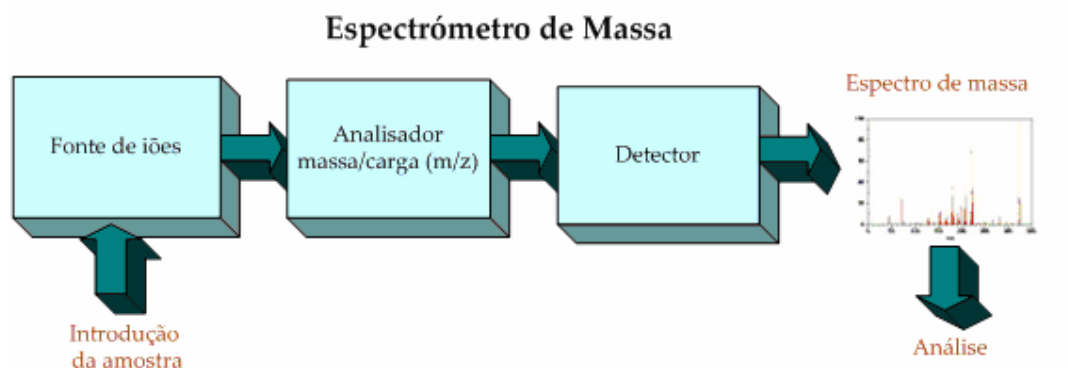


Figura 2 - Divisão do espectrómetro de massa em três componentes distintos, fonte de iões, analisador massa/carga e detector, obtendo por fim o espectro de massa da amostra analisada.

O espectro de massa é constituído por dois eixos cartesianos. No eixo dos xx fica representado o rácio (m/z), enquanto o eixo dos yy contém a percentagem da concentração relativa dos compostos detectados. O composto mais detectado ficará com 100%, enquanto os restantes compostos ficam com uma percentagem relativa ao composto mais detectado.

2.2.1 Introdução da amostra

O método de introdução da amostra na fonte de ionização depende do método de ionização a ser usado e da complexidade da mistura introduzida [1].

A amostra pode ser inserida directamente na fonte de ionização, ou submeter-se a algum tipo de cromatografia para a fonte de ionização [1].

Este último método de introdução da amostra envolve, normalmente, o acoplamento do espectrómetro de massa a um aparelho de cromatografia líquida (LC), cromatografia gasosa (GC) ou cromatografia líquida de alta performance (HPLC), entre outras [2].

Os cromatografos são classificados consoante as características das duas fases que os compõem, a fase móvel e a fase estática. A fase móvel é a estádio qual é misturada a amostra e a fase estática é a estádio no qual vai passar a fase móvel juntamente com a amostra. Há vários tipos de conjugação de fases, dependendo da necessidade da análise a realizar [6].

O LC utiliza uma fase móvel líquida constituída por uma solução de baixa viscosidade. Os componentes são primeiro dissolvidos num solvente e inseridos, juntamente com a fase móvel líquida, na coluna do cromatógrafo. A coluna do cromatógrafo contém uma fase estática sólida, a qual vai atrasar o avanço da fase móvel mediante a afinidade polar. Se introduzirmos a fase móvel sob pressão, iremos ter um HPLC, obtendo-se uma melhoria de análise em misturas quimicamente complexas [6].

O GC utiliza uma fase móvel gasosa constituída por um gás inerte, normalmente azoto ou hélio. A fase estática poderá ser constituída por um líquido ou sólido [2].

A amostra separada é então inserida, sequencialmente, no espectrómetro de massa para a análise.

2.2.2 Métodos de ionização da amostra

Muitos métodos estão disponíveis para a ionização mas cada método tem as suas vantagens e desvantagens. Os métodos de ionização que deverão ser usados para analisar a amostra dependem do tipo de amostra e do espectrómetro de massa disponível [2]. No entanto, os métodos mais usados no campo da bioquímica são a ionização por electrospray (ESI – *electro spray Ionisation*), desadsorção com laser assistida por matriz (MALDI – *matrix-assisted laser desorption ionization*) e o bombardeamento com átomos rápidos (FAB – *fast atom bombardment*) [1].

Em qualquer método de ionização é possível carregar as moléculas, positiva ou negativamente, consoante a afinidade dos protões presentes na amostra.

Posteriormente, o utilizador terá de decidir se vai detectar iões positivos ou negativos.

No modo de ionização positiva uma pequena quantidade de ácido fórmico é normalmente acrescentada à amostra, de forma a ajudar a protonação das moléculas. No modo de ionização negativa uma pequena quantidade de amónio, ou amido volátil, é acrescentada de modo a auxiliar a desprotonação das moléculas. Em todos os casos, a escala m/z deverá ser calibrada através de uma amostra padrão.

A ionização por ESI (Figura 3) é umas das técnicas de ionização à pressão atmosférica, e tem um espectro muito alargado de análise no que toca ao rácio massa/carga. É utilizada para detectar moléculas com pesos moleculares inferiores a 130 kDa com uma margem de erro na ordem dos 0.0001% [1].

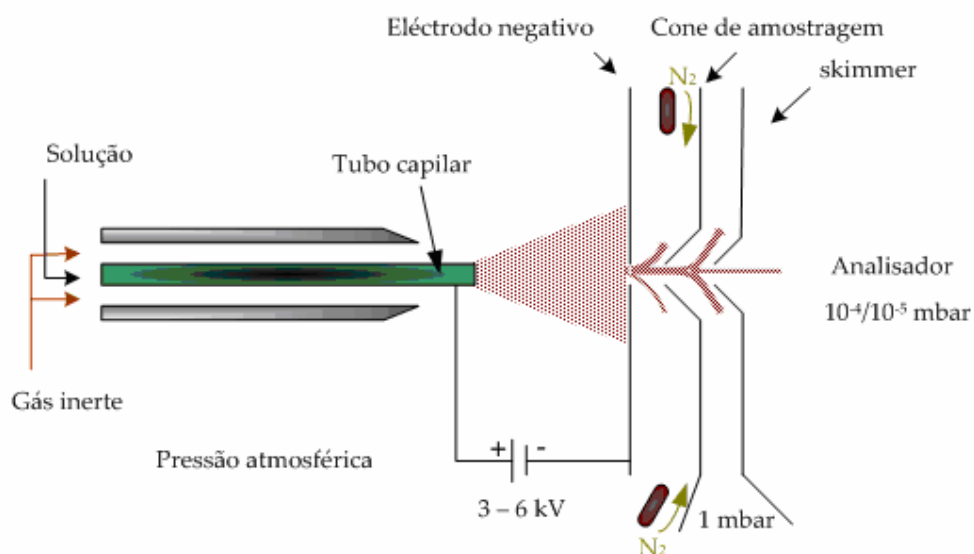


Figura 3 - Método de ionização por electrospray.

A amostra é primeiro dissolvida numa solução volátil polar e posteriormente bombeada para o tubo capilar (75 – 150 μ m), com um fluxo entre os 1 μ L/min e 1 mL/min. O tubo capilar está a um potencial entre 3 a 6 kV em relação ao electrodo contrário [7]. Em consequência deste enorme campo eléctrico a solução é transformada num aerosol de partículas altamente carregadas. A envolver o tubo capilar existe um outro tubo, pelo qual passa um gás inerte, normalmente azoto, que ajuda o aerosol a ser introduzido no espectrómetro de massa (Figura 4). As gotas carregadas diminuem de tamanho devido à vaporização do solvente, ajudado também pela presença do gás, conhecido por gás de secagem.

A pressão na região entre o tubo capilar e o electrodo é a pressão atmosférica. Os iões passam dessa região para uma região intermédia através de um orifício

no eléctrodo. Nesta região, os iões são submetidos a mais um banho de azoto passando para a região que fica compreendida entre o cone da amostra e o *skimmer*. Esta região intermédia é mantida a uma pressão mais baixa conseguida através de uma bomba rotatória. Os iões atravessam em seguida o *skimmer* em direcção ao analisador que se encontra a alto vácuo. O *skimmer* funciona como um separador deixando passar os iões mais pesados em direcção ao analisador, enquanto as moléculas mais leves de gás e solvente são bombeadas.

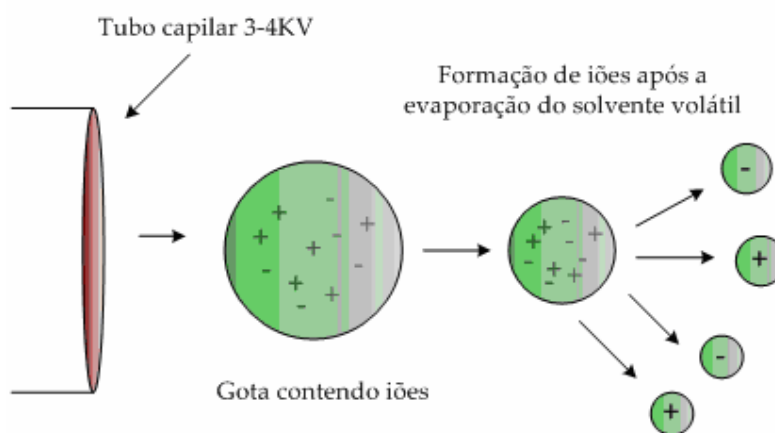


Figura 4 - Formação de iões pelo método de ionização por electrospray. Os iões isolados são formados através da evaporação do solvente volátil, no qual foi diluída a amostra.

A ionização através de MALDI é usada, essencialmente, em componentes orgânicos não voláteis, sensíveis à temperatura e com elevado peso molecular, na ordem dos 300 kDa [1].

É muito utilizado na área da bioquímica para analisar proteínas, peptídeos e oligo-nucleótidos [8].

MALDI é baseado no bombardeamento das moléculas em análise com um laser de baixa intensidade ionizando a amostra (Figura 5). A amostra é primeiro misturada numa matriz absorvente. Quando a matriz é atingida pelo laser, a energia deste é transformada em energia de excitação, provocando um espalhamento da matriz juntamente com a amostra [1].

MALDI é uma ionização suave e resulta, predominantemente, da geração de iões carregados com uma única carga [7, 9, 10].

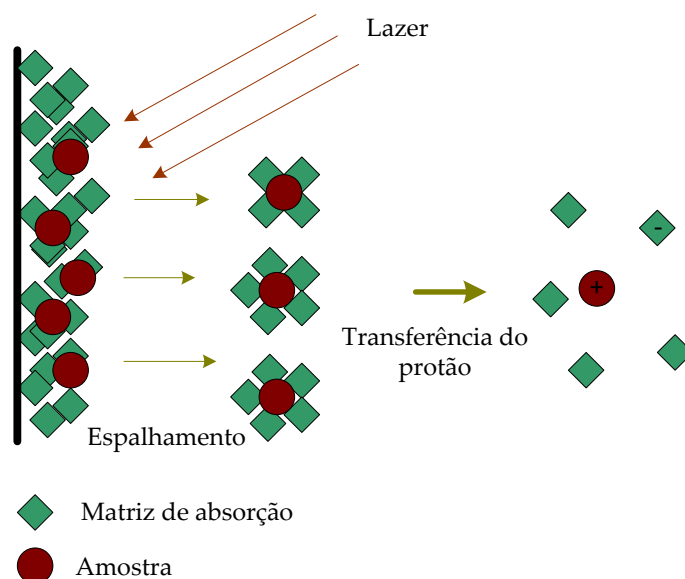


Figura 5 - Formação de íões através do método de ionização MALDI.

A ionização por FAB é muito similar à ionização através de MALDI. A solução para análise é misturada normalmente com glicerol sendo, posteriormente, submetida a um bombardeamento com partículas. Essas partículas são constituídas por um gás inerte (Ar ou Xe) com uma energia entre 4 a 10 KeV. O feixe de partículas ioniza a amostra e transfere-a para a fase gasosa.

É também uma ionização suave muito utilizada para amostras com baixo coeficiente volátil [1, 11].

2.2.3 Análise e separação de íões.

A função principal de um analisador de massas é separar os íões formados na fonte de ionização, de acordo com os seus rácios m/z .

Existem vários analisadores de massa disponíveis actualmente, como os quadrupolos, tempo de voo (TOF), sector magnético, quadrupolos *ion traps*, entre outros [12].

O quadrupolo (Figura 6) cria campos eléctricos, dentro do analisador, gerados por duas fontes de sinal de alternado. Os íões ao passarem pelos campos electro-magnéticos serão impelidos pelas forças geradas por estes. Dependendo da sua massa/carga, os íões conseguem progredir através do campo ou serão afastados de forma a não conseguirem passar pelo analisador [13].

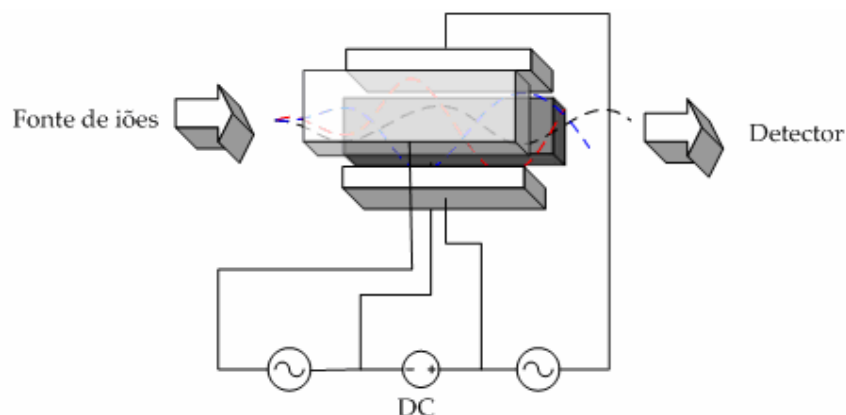


Figura 6 – Analisador quadrupolo. As partículas seguem trajectórias diferentes consoante a sua massa carga.

O analisador TOF simplesmente mede o tempo de voo que um ião demora desde a fonte de iões até atingir o detector (Figura 7).

Devido ao facto de todos os iões receberem a mesma energia cinética durante o tempo de aceleração, são separados em grupos consoante a sua velocidade (dependente da massa), percorrendo, assim, o campo aberto entre o gerador de iões até ao detector em diferentes períodos de tempo. Através da medição desse tempo é possível detectar o rácio massa/carga. Iões com menor carga percorrem a distância em menos tempo do que os iões com maior massa [1, 14].

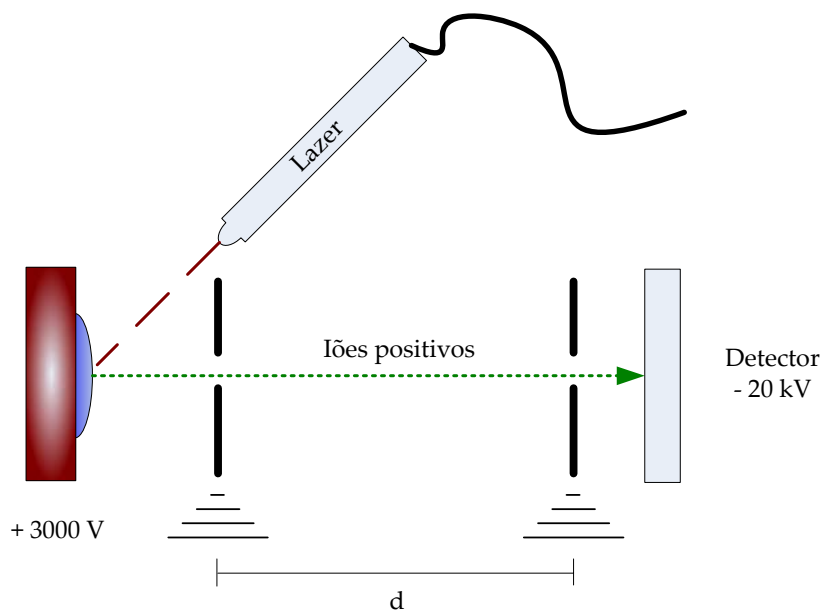


Figura 7 - Espectrómetro de massa TOF acoplado a um MALDI.

O TOF tem a vantagem de não impor limite de massa nas partículas a analisar, bem como a percentagem de iões que consegue detectar em relação aos outros métodos de espectrometria [14].

Os analisadores de massa têm diferentes características. Variam no rácio carga/massa que queremos analisar e na precisão que se pretende na análise.

Existem também diferentes compatibilidades com o método que pretendemos para a ionização. Por exemplo, todos os analisadores podem usar o ESI mas o quadrupolo não pode usar o MALDI [12].

2.2.4 Detecção de iões

O detector monitoriza as correntes de iões, amplifica e envia o sinal para um sistema computacional onde é gravado em forma de espectro de massa. O espectro de massa contém dois eixos, sendo o eixo dos xx para o rácio massa/carga e o eixo dos yy para a frequência relativa dos iões presentes na amostra.

A amplificação é necessária devido ao número reduzido de iões que chega ao detector num dado instante.

Existem vários tipos de detectores sendo o mais comum o multiplicador de electrões.

O multiplicador de electrões é um aparelho que detecta partículas eléctricas como electrões, fotões ou iões. A detecção da partícula é baseada em dois princípios. Primeiro, a partícula tem de ser convertida em electrões antes da amplificação. Este acontecimento ocorre quando o ião choca com a primeira placa (Figura 8). Posteriormente, esses electrões são amplificados em cascata. Os electrões são acelerados de forma a gerarem mais do que um electrão por electrão de embate nas placas. Se aumentarmos a diferença de potencial entre placas conseguimos um maior ganho na amplificação [1].

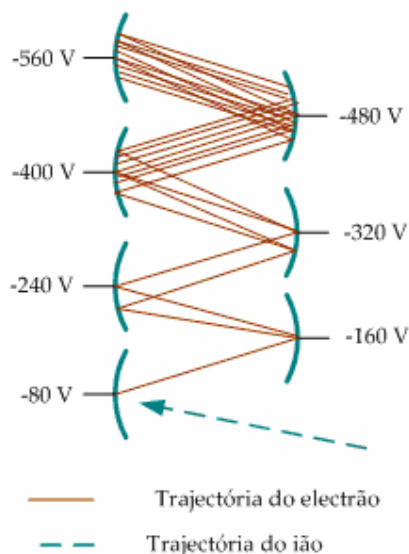


Figura 8 - Multiplicador de electrões.

2.3 Espectrómetros de massa em tandem

Os espectrómetros de massa em Tandem (MS/MS) são instrumentos constituídos por mais do que um analisador de massa. Existem no mercado dois, três ou quatro analisadores de massa acoplados. No entanto, os analisadores de massa não têm de ser do mesmo tipo. O MS/MS é composto por dois analisadores estando separados por uma câmara de colisão, a qual contém um gás inerte (árgon, xénon). O gás tem a função de colidir com as moléculas ionizadas seleccionadas no primeiro analisador, fragmentando-as.

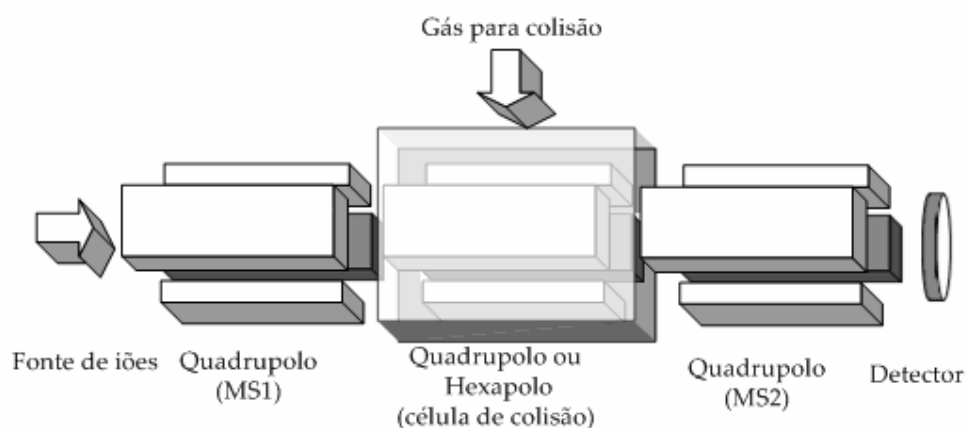


Figura 9 - Espectrómetro de massa MS/MS composto por dois quadrupolos e uma câmara de colisão.

Dois analisadores montados em tandem podem assumir as seguintes configurações:

- quadrupolo – quadrupolo;
- sector magnético – quadrupolo;
- sector magnético – sector magnético;
- quadrupolo – tempo de voo.

Os MS/MS são usados para obter informações estruturais das moléculas. O processo é simples. Fragmenta-se a molécula em iões dentro do espectrómetro de massa e analisam-se os resultados da fragmentação. A informação pode ser reunida num só espectro de massa gerando informação estrutural da molécula em questão. O espectrómetro de massa em tandem pode, também, detectar moléculas específicas em misturas complexas, baseadas nas suas características e nos padrões de fragmentação.

2.3.1 Analisar com o espectrómetro de massa em tandem

Existem quatro modos de aquisição de informação de forma a rastrear moléculas com os espectrómetros MS/MS [1, 15]:

- *Product ion scanning*;
- *Precursor ion scanning*;
- *Constant neutral loss*;
- *Select/Multiple Reaction Monitoring*.

Com o modo *Product ion scanning* (Figura 10), o MS1 é usado para seleccionar a passagem de iões com uma massa específica. Os iões seleccionados entram na câmara de colisão colidindo com o gás fragmentando-os. Os iões fragmentados serão posteriormente analisados pelo MS2.

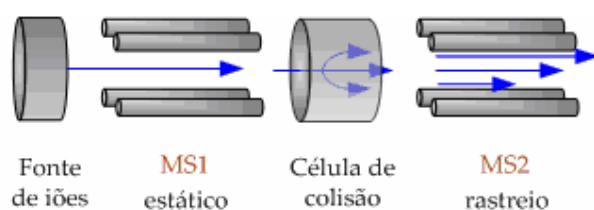


Figura 10 - Modo de análise através de *Product ion Scanning*.

Com o modo de *Precursor ion scanning* (Figura 11) o MS1 transmite todos os iões enquanto o segundo analisador é usado para filtrar iões específicos, originados pela colisão na câmara de gás.

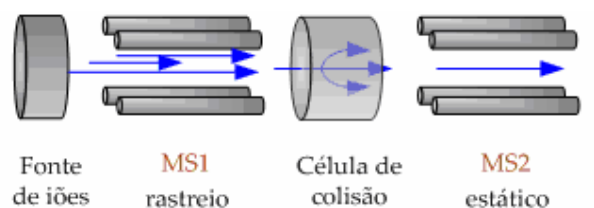


Figura 11 - Modo de análise através de *Precursor ion scanning*.

Com o método *Constant neutral loss* (Figura 12), é imposto aos dois espectrómetros que estejam em modo análise, cobrindo todo o espectro de massa/carga possível, tendo, o segundo, um desfasamento em relação ao primeiro, que difere num determinado número de unidades de massa em relação ao primeiro.

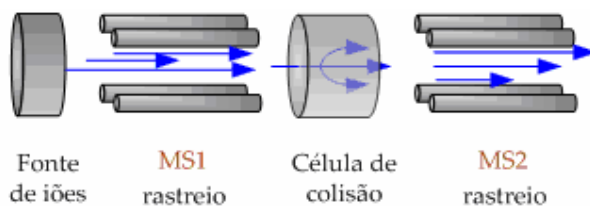


Figura 12 - Modo de análise através de *Constant neutral loss*.

Este método é usado para analisar os ácidos carboxílicos em soluções. O ácido carboxílico tende, normalmente, a fragmentar-se perdendo uma molécula neutra de dióxido de carbono, equivalente a 44Da. Todos os iões passam no primeiro espectrómetro em direcção à câmara de colisão devido a este varrer todo o espectro possível. Os iões que passam pelo segundo espectrómetro são aqueles que perderam 44Da na câmara de colisão.

O método *Multiple reaction monitoring* (Figura 13) impõe aos dois espectrómetros o modo estático, seleccionando, especificamente, uma determinada massa molecular.

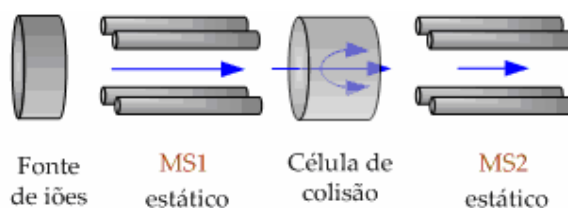


Figura 13 - Modo de análise através de *Multiple Reaction Monitoring*.

Este método só é utilizado quando o componente a ser rastreado tem características bem conhecidas. É usado, por exemplo, para efectuar testes de drogas através de amostras de sangue ou urina. Este método além de ser muito específico é também muito sensível.

2.4 Aplicações

Uma das primeiras áreas a beneficiar com a espectrometria de massa foi a indústria do petróleo. O petróleo é uma mistura muito complexa de moléculas, com uma composição muito variada dependendo da área geográfica onde é extraído. Esta variação química poderá trazer problemas aos especialistas operando nas centrais de refinamento e que convertem o petróleo numa panóplia de produtos utilizados no quotidiano. Foi a partir das décadas de quarenta e cinquenta, do século XX, que a espectroscopia de massa começou a ter grande impacto na indústria química [15]. Para perceber o porquê dessa importância é necessário demonstrar duas informações fundamentais que um

espectrómetro de massa fornece a um químico, nomeadamente, a massa das moléculas em análise e o seu espectro de massa.

A massa de uma molécula, claramente, a soma das massas dos átomos que a compõem. Diferentes elementos têm diferentes massas, então, por princípio, é possível distinguir as moléculas com massas aproximadas. Passemos a um pequeno exemplo para explicar melhor a informação obtida através do espectro de massa.

Três compostos químicos com massas por molécula muito aproximadas a 28 Da são o monóxido de carbono (CO), o nitrogénio (N₂) e o etileno (C₂H₄). Estes elementos encontram-se com facilidade na atmosfera. Em termos de massa atómica, o carbono representa-se por ¹²C, o azoto ¹⁴N, o oxigénio ¹⁶O e o hidrogénio ¹H. O número de cada elemento representa o número de prótons mais o número de neutrões de cada elemento, isto é, o número representa, aproximadamente, a massa do elemento em Daltons, mais precisamente 12.0000 para o carbono, 14.0067 para o azoto, 1.0079 para o hidrogénio e 15.9994 para o oxigénio. Esta nomenclatura é utilizada quando se torna necessário especificar um determinado isótopo. Então, as massas do CO, N₂ e C₂H₄ são 27.9994, 28.0134 e 28.0316 Da, respectivamente.

Se o espectrómetro de massa for suficientemente potente para poder separar os iões obtidos das moléculas e medir as suas massas, então, vamos conseguir identificar as moléculas presentes na amostra.

Na Figura 14, encontram-se os vários espectros de massa referentes às moléculas analisadas. No gráfico A, encontra-se o espectro de massa do monóxido de carbono o qual não se divide na câmara de colisão, devido às suas ligações triplas entre o carbono e o oxigénio. Na formação de iões fica com a forma de CO⁺, devido à perda de um electrão. Quando submetido a análise cria um espectro de massa com um pico de 100% na posição 28 [16].

No gráfico B, referente à análise do azoto, existem dois picos. O primeiro é referente à massa de um átomo de azoto, tendo o dobro da percentagem da molécula de azoto. Este comportamento fica-se a dever ao tipo de ligações que o átomo forma com outro átomo de azoto, sendo ligações muito electro-negativas [16]. Quando submetido à câmara de colisão existe uma percentagem de, aproximadamente, metade que sofre cisão, enquanto a restante permanece intacta. É devido a este facto que o pico de massa 14 tem o dobro da concentração do pico de massa 28.

No gráfico C, o etileno tem a particularidade de perder um átomo de hidrogénio no momento da ionização, não sofrendo qualquer cisão quando submetido à câmara de colisão. É por esta razão que o etileno aparece na posição com massa 27.

No gráfico D encontra-se o espectro de massa obtido com a análise feita aos três compostos em simultâneo, presentes numa solução em partes iguais. A linha referente à massa 28 aparece com 100%, enquanto as restantes massas têm a concentração a 66%.

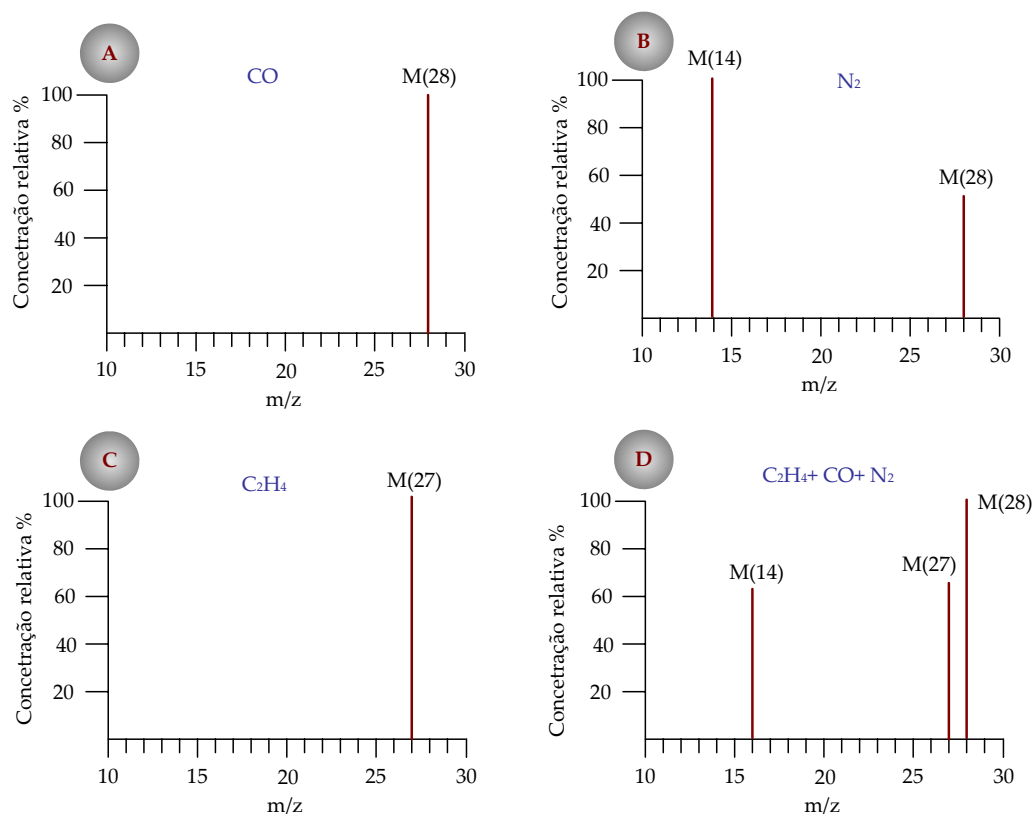


Figura 14 - Espectros de massa obtidos para cada composto isolado e se misturados todos os compostos com as mesmas concentrações.

Através da leitura dos espectros de massa o investigador consegue identificar os compostos presentes na amostra em análise. No entanto, quando se analisam amostras quimicamente complexas, o investigador terá de ter uma certa sensibilidade/experiência para ler os espectros de massa, pois a quantidade de informação poderá ser enorme, como se deixa adivinhar pelo exemplo.

2.5 Sumário

Os espectrómetros de massa evoluíram muito desde o seu aparecimento, embora os princípios básicos ainda se mantenham. Actualmente existem variados tipos de espectrómetros como os quadrupolos, o TOF ou o sector magnético. Mas o melhoramento dos resultados da análise não se centra

somente nos avanços dos espectrómetros, mas também no melhoramento das técnicas de ionização como o ESI e o MALDI. A junção de vários espectrómetros de massa em tandem veio trazer melhoramentos em termos de desempenho, principalmente na obtenção de informação sobre as partes estruturais das moléculas.

Com o recurso a esta tecnologia é possível com uma simples análise identificar os elementos presentes em amostras quimicamente complexas num curto espaço de tempo.

3 Conceitos biológicos

Os processos biológicos envolvidos nas formas de vida são de tal forma complexos, que qualquer pequena alteração dos processos metabólicos poderá trazer anomalias para o organismo. No entanto, são também algumas destas anomalias que trazem vantagem competitiva a um determinado organismo, para que a selecção natural continue o seu caminho.

3.1 Células

Uma “célula” é a mais pequena porção do organismo capaz de vida independente [17]. Todos os seres vivos compõem-se de uma, ou mais células. Alguns seres microscópicos são compostos por uma única célula, como as bactérias, mas os animais e plantas são formados por muitos milhões de células organizadas em tecidos e órgãos.

Apesar da grande variedade de seres vivos, as células que os compõem têm muito em comum. Existem dois tipos principais de células, as procarióticas e eucarióticas. As procarióticas possuem um único compartimento encerrado numa membrana e não possuem quaisquer estruturas ou compartimentos bem definidos no citoplasma, material gelatinoso contido no interior da membrana plasmática (celular). Exemplos de organismos procarióticas são as Eubactérias e Archeobactérias. Os organismos vivos possuem células eucarióticas, que contêm um certo número de compartimentos especializados (Figura 15). Estes compartimentos são envolvidos por membranas e têm o nome de organitos. Os organitos criam uma série de microambientes dentro da célula, nos quais certas sequências de reacções se podem processar com a máxima eficiência [18]. Exemplos de organismos eucarióticas são as plantas, fungos e animais.

O organito mais importante das células eucarióticas é o núcleo, o qual encerra o material genético. É o núcleo que contém informação codificada em moléculas de ácido desoxirribonucleico (ADN), sendo esta informação a regular a actividade da célula e a assegurar a sua duplicação e transmissão das suas características à descendência.

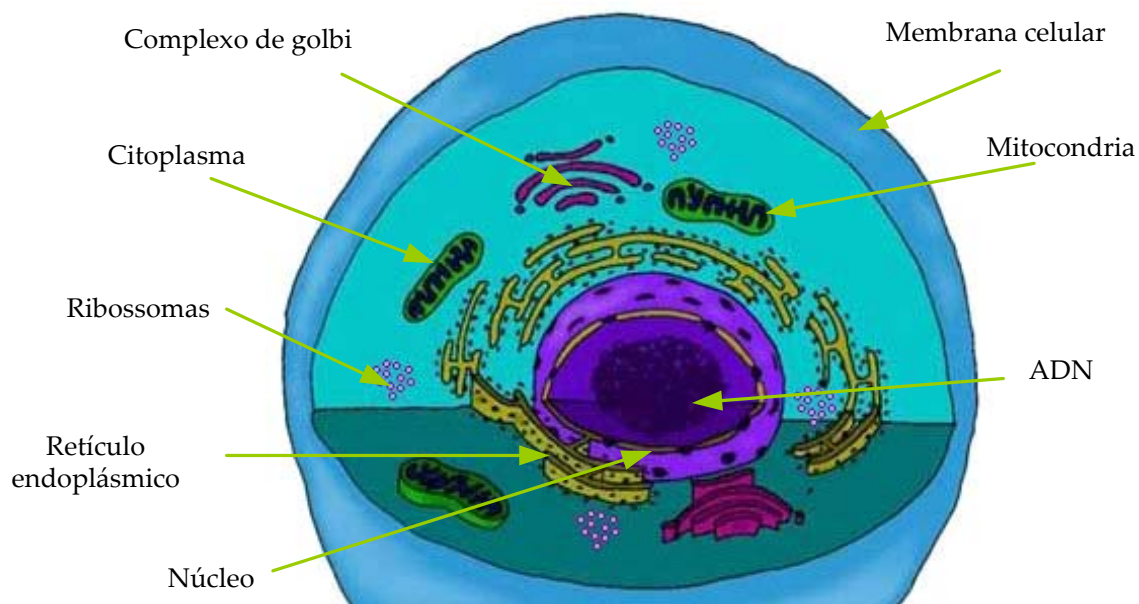


Figura 15 – Visão geral de uma célula eucariótica.

Um dos outros organelos semi-autónomos das células eucarióticas é a mitocôndria, responsável pela produção de energia durante a respiração aeróbica (processo responsável pela formação da molécula trifosfato de adenosina, ATP, a qual contém uma ligação de fosfato de alta energia). Basicamente, transformam a energia dos alimentos em energia biologicamente útil. A mitocôndria contém o seu próprio código genético e novas mitocôndrias surgem por divisão ou fragmentação das mitocôndrias existentes. Existem duas formas de produzir ATP na mitocôndria. Na primeira forma a mitocôndria recorre aos glícidos e hidratos de carbono para produzir ATP, mas quando detecta que esses níveis passam abaixo de um determinado limiar recorre aos ácidos gordos. Normalmente esta fase ocorre quando o organismo executa uma elevada actividade física. É nesta fase de produção de ATP que a mitocôndria necessita de ácidos gordos de cadeia longa que são transportados para a mitocôndria por uma molécula que tem por nome carnitina. A carnitina é responsável pelo transporte de ácidos gordos de cadeia longa para o interior da mitocôndria, e também é responsável pela extracção de ácidos gordos de cadeia curta do interior da mitocôndria. Os ácidos gordos de cadeia curta são o resultado da transformação de ácidos gordos de cadeia longa juntamente com oxigénio em ATP. A acilcarnitina é a molécula que resulta da junção da carnitina com uma molécula de ácidos gordos [18].

O ribossoma, outro dos organelos presentes na célula, é a entidade responsável pela sintetização de proteínas. Cada ribossoma consiste em duas subunidades, a grande e a pequena.

Existem muitos outros organitos, mas não têm interesse relevante para o estudo que se segue.

3.2 Biologia molecular

O genoma é o conjunto de todos os genes presentes num organismo contendo toda a informação necessária para expressar proteínas necessárias ao normal funcionamento do mesmo. O genoma está organizado em cromossomas, podendo os organismos ter mais do que um cromossoma. Os humanos possuem 24 pares de cromossomas distintos, enquanto as bactérias possuem apenas um cromossoma. Cada cromossoma é um par de moléculas de ADN, ligadas por quatro bases distintas: A, C, G e T (adenina, citosina, guanina e timina), também chamados nucleótidos (Figura 16) [19].

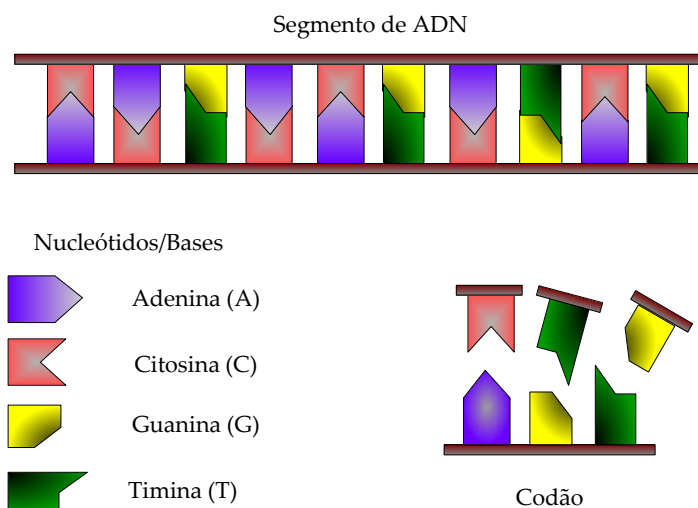


Figura 16 - Exemplo de um segmento de ADN. A adenina só emparelha com a citosina e a guanina só emparelha com a timina.

Os cromossomas têm partes codificantes e partes não codificantes. Os genes são as partes codificantes contendo os códigos para produzir proteínas ou cadeias polipeptídicas. O gene especifica a sequência de aminoácidos de uma proteína ou de uma das cadeias polipeptídicas (Figura 18), ou seja, os genes definem as regras de conversão da informação que é escrita, com base num alfabeto de quatro letras (A, C, T, G), através de um alfabeto de 20 letras tendo o nome de aminoácidos presentes no anexo 1.

Existem mais de 200 aminoácidos presentes na natureza, mas normalmente são necessários apenas 20 para sintetizar as proteínas necessárias ao metabolismo das células [19].

Os organismos heterótrofos, como é o caso dos seres humanos, não conseguem sintetizar alguns dos aminoácidos necessários à produção das cadeias polipeptídicas. Esses aminoácidos, denominados os aminoácidos essenciais, são obtidos a partir dos alimentos ingeridos pelos organismos, sendo eles leucina, isoleucina, valina, triptofano, metionina, fenilalanina, treonina e lisina. Os restantes, conhecidos como aminoácidos não-essenciais são produzidos pelo organismo. Existe uma exceção com a histidina sendo um aminoácido essencial na infância.

Para fazer a correspondência entre as bases e os aminoácidos, as bases agrupam-se três a três formando codões. Cada codão corresponde a um aminoácido, mas um aminoácido pode corresponder a mais do que um codão. As quatro bases combinadas, três a três, dão 64 codões possíveis, mais do que o necessário para codificar os 20 aminoácidos diferentes. Diz-se que o código é “degenerado” porque alguns dos aminoácidos são especificados por mais do que um codão (Figura 17). Alguns codões desempenham outros papéis, como, por exemplo, codões de “iniciação” (ATG) e de “finalização” (TAA, TGA e TAG) para assinalar o começo e o fim de um gene.

TTT TTC	Phe	TCT TCC	Ser	TAT TAC	Tyr	TGT TGC	Cys
TTA TTG	Leu	TCA TCG		TAA TAG	Stop	TGA TGG	Stop Trp
CTT CTC CTA CTG	Leu	CCT CCC CCA CCG	Pro	CAT CAC	His	CGT CGC CGA CGG	Arg
ATT ATC ATA	Ile	ACT ACC ACA ACG	Thr	AAT AAC	Asn	AGT AGC	Ser
ATG	Met			AAA AAG	Lys	AGA AGG	Arg
GTT GTC GTA GTG	Val	GCT GCC GCA GCG	Ala	GAU GAC	Asp	GGU GGC GGA GGG	Gly
				GAA GAG	Glu		

Figura 17 - Tabela de correspondência de codões para aminoácidos.

O gene é traduzido em proteína por um processo que envolve duas fases, a transcrição e a tradução (Figura 18). No processo de transcrição, o ADN do gene é utilizado como molde para a síntese de uma cadeia complementar. Posteriormente, os intrões, sequências de bases não codificantes presentes no

gene, são removidos da cadeia por meio de enzimas num processo chamado *splicing*. Obtém-se uma molécula de ácido ribonucleico mensageiro (ARNm) que se encaminha para o exterior do núcleo através dos seus poros, dando início ao processo de tradução de proteínas. O ARNm também possui, como é óbvio, as quatro bases distintas, mas com uma diferença, a timina (T) foi substituída pelo uracil (U). A molécula ARNm contém o código para expressar uma proteína específica.

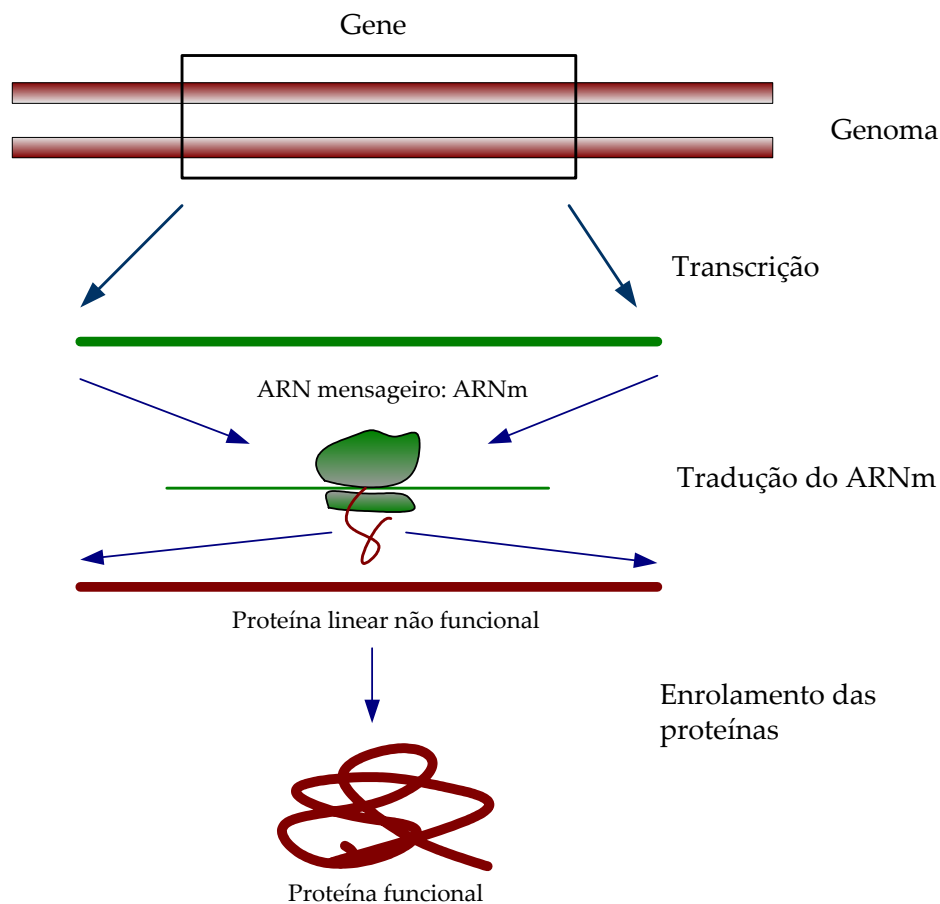


Figura 18 – Todo o processo de transformação de um gene em proteína.

O processo de tradução é realizado no citoplasma, no qual os ribossomas são unidos à molécula de ARNm codificando a estrutura primária de uma proteína específica (Figura 19).

Cada ribossoma compõe-se de duas sub-unidades, uma pequena e uma maior. Quando se combinam para formar o ribossoma proporcionam o meio certo para a síntese das proteínas, assim como os locais específicos de ligação para os portadores de aminoácidos e para a cadeia de ARNm. Os aminoácidos são transportados para o ribossoma por uma outra forma de ARN, o ARN de transferência (ARNt). Há pelo menos um tipo de molécula de ARNt para cada

tipo de aminoácido. Cada molécula de ARNt apresenta uma sequência de três bases chamada de anti-codão. Estas emparelham com as bases dos codões de ARNm para codificar o aminoácido transportado por cada ARNt. Assim, o ARNt funciona como um adaptador, trazendo os aminoácidos correctos especificados pelos codões do ARNm.

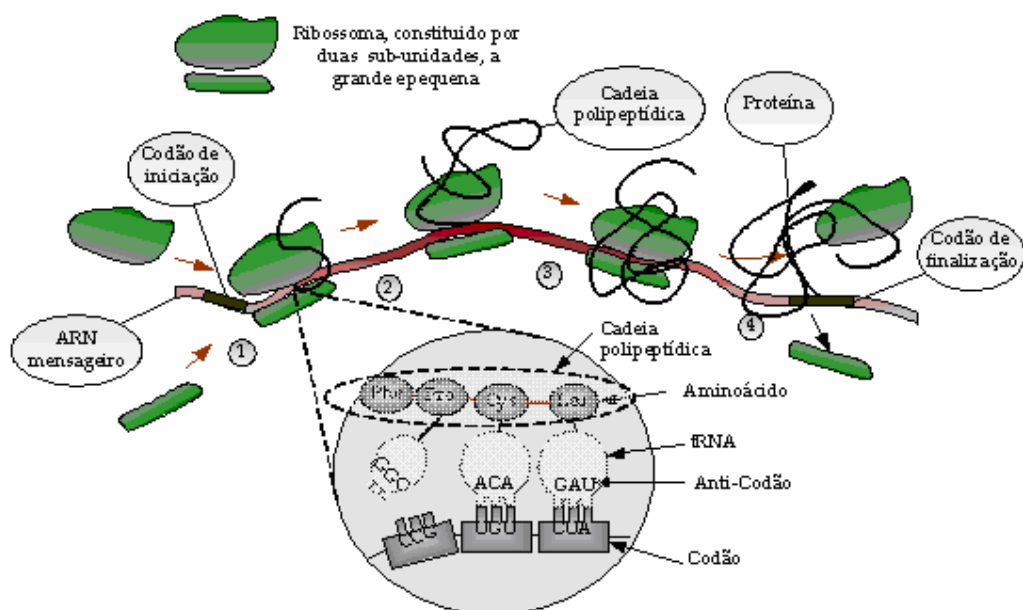


Figura 19 - Processo de tradução do ARN mensageiro.

À medida que cada aminoácido é trazido para o ribossoma pelo seu ARNt, uma enzima ribossómica cataliza uma ligação peptídica entre ele e o aminoácido ligado ao ARNt precedente, construindo-se assim uma cadeia polipeptídica. A tradução prossegue até se chegar a um codão de finalização; nessa altura, a cadeia polipeptídica é libertada do ribossoma e as duas subunidades separam-se. Uma só molécula de ARNm pode ter mais de um ribossoma a deslocar-se ao longo dela num dado momento, formando assim mais do que uma cadeia polipeptídica. Posteriormente, procede-se ao enrolamento da cadeia polipeptídica/proteína, a qual passa a proteína funcional.

Existem vários tipos de proteínas estando agrupadas pelas diversas funções a elas atribuídas, como mostra a Tabela 1.

Tabela 1 - Lista dos vários grupos de proteínas existentes.

Grupo	Função	Exemplos
Enzimas	Funções catalíticas.	As células do fígado contêm milhares de enzimas, cada uma responsável para catalizar de determinada reacção. Transformar o aminoácido de fenilalanina em tirosina.
Estruturais	Providencia mecanismos de suporte para as células e tecidos.	As células exteriores como pele, cabelo, etc.; as células interiores, como os tendões.
Transporte	Transporta pequenas moléculas ou iões.	No sangue a hemoglobina transporta o sangue, <i>albumin</i> transporta o lípidos, entre outros.
Motoras	Gera movimento nas células e tecidos.	O <i>myosin</i> providencia a força para os movimentos nos animais.
Armazenamento	Armazena pequenas moléculas e iões.	O ferro é armazenado no fígado através da proteína <i>ferritin</i> ; <i>ovalbumin</i> na clara do ovo é usado como fonte de aminoácidos para o desenvolvimento das aves embrionárias.
Sinais	Transporta sinais de célula para célula.	Muitas das hormonas que regulam os factores de crescimento são proteínas; a insulina é uma proteína que controla os níveis de glucose no sangue.
Receptoras	Usado pelas células para detectar sinais e transmiti-los às células responsáveis pelo seu processamento.	A <i>rhodopsin</i> na retina detecta a luz.
Regulação genética	Ligada ao ADN liga ou desliga a expressão de determinado gene.	A lactose repressora na bactéria silencia o gene que codifica a enzima que degrada a <i>lactose sugar</i> .
Funções especiais	Muito variadas.	Os organismos desenvolvem proteínas com aplicações muito específicas. A <i>antifreeze protein</i> dos peixes do Ártico e Antártico protegem do congelamento do seu sangue.

O tamanho de uma proteína pode variar entre 30 a 10.000 aminoácidos, mas normalmente, está compreendido entre 50 a 2000 aminoácidos [18].

3.3 Doenças Metabólicas

As doenças metabólicas são causadas por erros inatos do metabolismo e são estudadas por uma área específica da genética, a bioquímica genética. Para algumas já há tratamento dependendo da doença, da habilidade do clínico e da família da criança. Infelizmente para a maioria, ainda não se dispõe de terapêutica eficaz. Neste caso, o tratamento restringe-se em oferecer à criança suporte clínico e medidas que visam melhor qualidade de vida.

A identificação de cada doença é realizada, não pelas suas características clínicas, mas pela investigação laboratorial que permite o diagnóstico através da

identificação das alterações bioquímicas presentes. Os subgrupos são vários e o número de desordens conhecidas ronda as duzentas [20]. A clínica é bastante variável e a maioria dessas doenças é de incidência rara, o que traz dificuldades para o diagnóstico.

Estas doenças são caracterizadas pela produção anormal de proteínas, em excesso ou em defeito, o que terá uma implicação directa na redução ou anulação da acção catalítica presente na célula. Se determinada enzima, importante para uma actividade metabólica, for deficitária irá produzir em défice uma determinada molécula A, indo aumentar a concentração da molécula B que estava na origem da molécula A. Como a célula tem tendência a repor o equilíbrio enzimático, irá utilizar um método alternativo, ou seja diminuir a concentração da molécula B. O resultado será a produção de uma molécula C, que possivelmente em elevadas concentrações será tóxica para a célula, podendo danificá-la ou inibir outras reacções metabólicas. Devido ao método alternativo de metabolismo ser menos eficiente do que o principal, poderá ocorrer um aumento de concentração da molécula B podendo também ser prejudicial para a célula.

As doenças rastreáveis através da técnica de MS/MS agrupam-se em categorias consoante os marcadores que as definem e os processos metabólicos em que os respectivos marcadores estão envolvidos. No caso das Aminoacidopatias os marcadores que estão na sua origem são os aminoácidos, existindo deficiências várias no metabolismo dos mesmos [21]. As Acidurias Orgânicas abrangem as deficiências metabólicas relativas à interacção dos aminoácidos com ácidos gordos e carbo-hidratos [21]. Este grupo de doenças caracteriza-se por elevadas concentrações de ácidos orgânicos no sangue e urina, estando presentes em variados processos intermédios do metabolismo [20]. O grupo das doenças hereditárias da β -oxidação mitocondrial dos ácidos gordos caracteriza-se por uma elevada concentração de acilcarnitinas de cadeia longa, devido a problemas de oxidação dos mesmos, dentro do organito da mitocôndria [20].

A Tabela 2 contém o nome das doenças que farão parte do rastreio nacional, juntamente com os marcadores que terão de ser observados para cada doença [22]. O recém-nascido é possuidor de determinada doença se registar uma alta ou baixa concentração de moléculas a que se refere o marcador.

Tabela 2 - Doenças rastreáveis em Portugal recorrendo ao (MS/MS). Os marcadores correspondem aos aminoácidos ou acilcarnitinas a ter em conta para as referidas doenças. As quatro primeiras linhas dos marcadores correspondem aos aminoácidos, que estão representados por três letras conforme convenção já estabelecida. As restantes linhas corresponderem às acilcarnitinas que estão sob a forma *L-isomeric*. Esta forma indica que C corresponde a uma acilcarnitina seguido do número de carbonos que compreende o grupo do ácido gordo. O OH representa a presença de um grupo de ácidos gordos hidroxilados. A presença de uma ligação dupla entre carbonos (saturação) está representada por dois pontos seguido da posição. O DC corresponde a ácidos dicarboxílicos.

Doenças	Marcadores de doença
Aminoacidopatias	
Fenilcetonúria (PKU)	Phe, Phe/Tyr
Leucínose (MSUD)	Leu/Ile, Val
Citrulinemia (CIT)	Cit
Acidúria ArgininoSuccínica (ASA)	Cit
Acidúrias Orgânicas	
Acidúria Propiónica (PA)	C3
Acidúria Metilmalónica (MAA)	C3
Acidúria Isovalérica (IVA)	C5
Acidúria 3-Hidroxi-3-MetilGlutárica (3-HMG)	C5-OH
Acidúria Glutárica Tipo I (AG I)	C5DC
Doenças hereditárias da β-oxidação mitocondrial dos ácidos gordos	
Deficiência da Desidrogenase dos Ácidos Gordos de Cadeia Média (MCAD)	C8, C10, C10:1, C6
Deficiência da Desidrogenase dos Ácidos Gordos de Cadeia Muito Longa (VLCAD)	C14:1, C14, C16
Deficiência da Desidrogenase de 3- Hidroxi-AcilCoA de Cadeia Longa (LCHAD)	C16-OH, C18:1-OH, C18-OH
Deficiências em Carnitina-Palmitoil Transferase I (CPT I) e Carnitina-Palmitoil Transferase II (CPT II)	C16, C18:1, C18

A descrição mais pormenorizada das doenças encontra-se no anexo 2.

Uma doença que está incluída no rastreio neonatal e que não recorre ao MS/MS para seu diagnóstico é o Hipotiroidismo Congénito (HC). A sua origem é hereditária ou esporádica, causando uma diminuição da hormona da tiróide (T4) presente no sangue. Esta hormona é responsável pela regulação do funcionamento do organismo e é segregada pela glândula da tiróide. Quando existe uma diminuição de T4 a glândula pituitária segrega a hormona de estimulação da tiróide (TSH), enviando assim um sinal para a glândula da tiróide segregar mais T4.

As causas desta patologia são o atraso de crescimento, atraso no desenvolvimento do processo cognitivo e eventualmente provoca deficiência mental. O tratamento consiste em administrar ao paciente a hormona T4 [23].

É utilizada a técnica da imunofluorescência que consiste em utilizar anticorpos ligados quimicamente a uma molécula fluorescente com afinidade para criar ligações com outras moléculas específicas em amostras. Neste caso a molécula será a TSH medindo-se posteriormente a luminosidade da análise.

Existe uma discussão em torno da detecção do HC. Nos Estados Unidos o rastreio passa por medir os níveis de T4 presentes no sangue dos indivíduos, enquanto na Europa o rastreio é efectuado medindo a concentração do TSH presente no sangue. Embora o teste do TSH seja mais específico, o teste do T4 consegue ser mais sensível na detecção, mas menos específico na detecção de duas variantes do HC [24].

Na Tabela 3 encontram-se o número de casos e suas respectivas frequências, detectados em Portugal no ano de 2002, 2003 e 2004 [25, 26].

Tabela 3 – Número de casos detectados em 2002, 2003 e 2004

Ano	Recém-nascidos estudados	Doença	Nº de casos	Frequência
2002	113.897	HC	37	1 / 3078
		PKU	16	1 / 7118
2003	112.557	HC	41	1/2745
		PKU	9	1/12506
2004	108.564	HC	43	1/2524
		PKU	11	1/9869

3.4 Sumário

É nas células que ocorrem os processos biológicos necessários para manter viva toda a estrutura de um ser vivo. Quando algo se descontrola e deixa de funcionar correctamente a célula tem mecanismos para compensar essas falhas. Mas nem sempre é possível a compensação, pois para equilibrar um determinado processo metabólico pode-se desequilibrar outro. Por vezes, o desequilíbrio é de tal forma, que a célula não tem capacidade para o corrigir. No entanto, é possível a detecção de desequilíbrios metabólicos através da quantificação dos compostos envolvidos no metabolismo. Com este processo pode-se facilmente rastrear os recém-nascidos, de forma a detectar falhas que darão origem a doenças antes de surgirem os seus sinais aparentes.

Esta solução oferece grandes vantagens para os médicos, pois é possível efectuar a correcção dos níveis de aminoácidos ou acilcarnitinas de forma a minorar os danos para o recém-nascido.

4 Rastreio de doenças em recém-nascidos

Os primeiros dias de vida de um recém-nascido podem ser críticos para o seu bem-estar futuro, porque pode ser o único momento em que os especialistas têm a possibilidade de diagnosticar e tratar uma determinada doença metabólica. Para tal, é necessário recorrer a um conjunto de tecnologias para proceder ao rastreio de doenças do foro metabólico.

Com recurso a uma tecnologia já muito experimentada noutros campos, o MS/MS, é possível de uma forma rápida e sustentável diagnosticar uma elevada bateria de patologias. Contudo, para algumas patologias ainda não existe uma possibilidade de tratamento. Este problema levanta uma grande questão de ordem ética. Devem ou não os pais tomarem conhecimento que o seu recém-nascido é possuidor de uma doença de carácter genético sem possibilidade de tratamento?

A resposta a esta questão de grande importância está, no entanto, fora do contexto deste trabalho. Aqui procurámos apenas fornecer aos técnicos os melhores meios para que o diagnóstico seja realizado em tempo útil.

4.1 Recursos (tecnologia, ferramentas, sistemas de informação...)

Nos meados dos anos 30 foi reconhecido que uma das causas de morte prematura, ou atrasos variados no desenvolvimento dos recém-nascidos, era devido a desordens do metabolismo de aminoácidos. Posteriormente, verificou-se que também a existência de desordens na oxidação dos ácidos gordos, originavam sintomas semelhantes [22, 27].

Mas foi em 1961 que, Robert Guthrie e a sua equipa, desenvolveram um teste de forma a detectar a presença de níveis elevados do aminoácido Fenilalanina, no sangue dos recém-nascidos. A recolha de amostras de sangue eram realizadas em papel absorvente, submetendo a amostra recolhida a um teste com bactérias que ele próprio idealizou [28]. Através da inibição do crescimento da bactéria *Bacillus subtilis*, era testada a presença de níveis elevados do aminoácido Fenilalanina no sangue dos neo-natais, começando, assim, o diagnóstico precoce de Fenilcetonúria (PKU). Este teste tinha o inconveniente de originar um elevado número de falsos positivos [21].

Em meados dos anos 80, Chalmers demonstrou o valor terapêutico ao medir a concentração de ácidos gordos [29], propondo, mesmo, uma solução para a regulação das deficiências [30] relativas às doenças metabólicas com origem nos ácidos gordos. No entanto, já em 1977, Bieber tentava medir e isolar acilcarnitinas [31]. Este tipo de análises recorriam à cromatografia gasosa analisando amostras de urina [32].

Posteriormente decorreram vários testes para analisar os ácidos gordos e acilcarnitinas com aproximações preliminares aos perfis das doenças metabólicas, os quais já recorriam a um GC/MS (cromatografia gasosa acoplado a um espectro de massa) para analisar as amostras de urina [32-35].

Com a introdução de técnicas de ionização através do FAB para o MS, os compostos com elevada polaridade, como por exemplo as acilcarnitinas, vão poder ser analisadas de uma maneira simples e rápida [36].

Paralelamente ao desenvolvimento de técnicas de análise de compostos, vão sendo melhoradas as aptidões de diagnóstico permitindo, assim, um melhor reconhecimento das patologias devido aos erros de metabolismo [37].

Na década de 90, com a introdução da espectrometria de massa em tandem (MS/MS), foi possível melhorar a análise das acilcarnitinas presentes nas amostras de urina e dos aminoácidos presentes nas amostras de sangue, sem necessidade de qualquer forma de cromatografia [38, 39]. Posteriormente, a metodologia foi melhorada e recombinação, necessitando de analisar somente a amostra de sangue para obter os níveis de acilcarnitinas e aminoácidos presentes num determinado indivíduo [40-45]. Com um único ensaio, aproximadamente de 2 minutos, os níveis de aminoácidos e acilcarnitinas, presentes na amostra de sangue do recém-nascido, são obtidos. Tem ainda a vantagem de se poder aumentar o número de marcadores a analisar sem ter a necessidade de alterar a tecnologia e os protocolos existentes, possibilitando assim aumentar o número de doenças a rastrear [46].

A metodologia que recorre ao MS/MS continua praticamente inalterada até aos dias de hoje.

Todo o procedimento que começa pela recolha de amostras de sangue e termina na obtenção dos valores das intensidades de aminoácidos e acilcarnitinas, é um processo que deverá seguir as recomendações do *Center for Disease Control and Prevention* (CDC) [47]. O centro está sediado em Atlanta/Georgia sendo uma agência do departamento da saúde e serviços humanos dos Estados Unidos da América. A agência fornece um conjunto de relatórios e meios técnicos de forma a prestar auxílio aos vários centros de rastreio nacionais e internacionais, como é o caso do Instituto de Genética Médica Abel Salazar no Porto (IGM).

A recolha de sangue é efectuada no recém-nascido entre o primeiro e o segundo dia de vida, sendo recolhido através de um cartão absorvente [48] (Figura 20). Devido ao elevado stress e dor que o recém-nascido fica sujeito no momento da recolha de sangue, é, normalmente, utilizado um creme para anestesiar o local onde a picada se vai proceder. O creme utilizado tem por nome *Eutectic Mixture of Local Anesthetics* (ELMA), trazendo inconvenientes para o especialista no momento da análise dos resultados. Os resíduos de creme poderão dificultar a medição dos níveis de acilcarnitinas perturbando o trabalho do especialista em determinar as possíveis patologias do recém-nascido [49]. Posteriormente, a amostra é seca e enviada para o laboratório de análise.

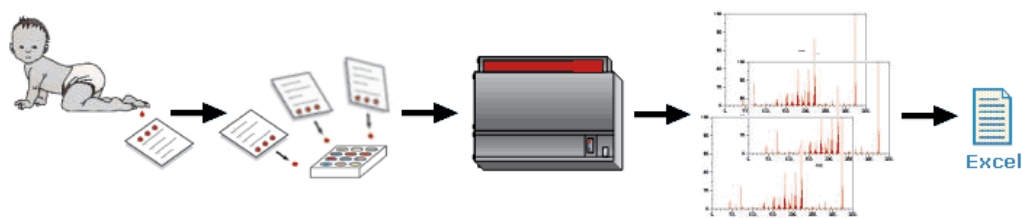


Figura 20 – Processo de rastreio dos recém-nascidos.

No laboratório o cartão é perfurado no local onde o sangue foi absorvido sendo posteriormente extraído por meio de solventes (metanol). À solução obtida adicionam-se os padrões referência para os diferentes aminoácidos ou acilcarnitinas a analisar. Os padrões de referência são isótopos dos aminoácidos e acilcarnitinas para não existir sobreposição dos níveis a analisar [22]. Estes padrões de referência permitem a quantificação dos aminoácidos e acilcarnitinas em análise [21].

Juntamente com os padrões de referência também são inseridos na amostra dos indivíduos, padrões de intensidade sugeridos pelo CDC. Os padrões de intensidade, quando analisados pelo espectrómetro, terão de estar acima de um determinado limite. Caso contrário, terá de se repetir a análise para o indivíduo que corresponde ao poço [22, 50]. Este controlo serve para verificar se a solução foi aspirada em quantidade suficiente para ser analisada pelo MS/MS.

As soluções dos diversos recém-nascidos são inseridas em diferentes poços de uma placa, que será introduzida num robot que automatiza o processo de aspiração da solução para o interior do MS/MS. A placa é constituída por 96 poços, sendo os primeiros preenchidos com soluções padrões fornecidas pelo CDC, chamados controlos. No caso do IGM são quatro os controlos, mas a recomendação do CDC especifica mais do que dois. Os controlos têm presente um número específico de aminoácidos e acilcarnitinas, assim como uma concertação específica. Quando analisadas, as concentrações terão que estar

compreendidos num determinado limite. Caso contrário terá que se realizar nova preparação da placa [22, 50].

Todos os padrões de referência e controlo são cedidos gratuitamente aos laboratórios por parte do CDC. Em troca, os laboratórios terão de fornecer um conjunto de dados referentes aos níveis de concentrações de aminoácidos e acilcarnitinas obtidos aquando do rastreio dos recém-nascidos. Estes dados serão analisados pelo CDC para produção de relatórios publicados anualmente [50], de forma a melhorar a qualidade de rastreio nos laboratórios através do programa *Newborn Screening Quality Assurance Program* (NSQAP).

Os aminoácidos e as acilcarnitinas são detectados pelo MS/MS recorrendo a técnicas distintas.

Os aminoácidos partilham um mesmo grupo químico formado pelo ácido fórmico (CH_2O_2) ao qual se recorre para proceder à análise através do MS/MS. Quando os aminoácidos são sujeitos à câmara de colisão produzem um fragmento de 42Da que corresponde ao ião ácido fórmico [21]. Devido à necessidade de melhorar a eficiência de ionização e aumentar a sensibilidade da espectroscopia de massa adicionam-se butilos (C_4H_9) à amostra [20]. O butilo vai-se ligar ao grupo carboxilo (CO_2H), o qual se vai separar do aminoácido quando submetido à câmara de colisão. Este irá formar um fragmento de 102Da não iónico. Assim, o MS/MS, no modo de operação *constant neutral loss* NL-102, permite detectar vários aminoácidos. Partilham em comum este tipo de fragmentação os aminoácidos fenilalanina, tirosina, leucina, alanina, valina e metionina [21].

Existe um sub-grupo de aminoácidos básicos como a ornitina, citrulina e homocitrulina que, quando submetido à câmara de colisão, irá formar partículas que contêm o butilo juntamente com o grupo carboxilo 44Da e mais o grupo amónia 17Da. Este conjunto perfaz uma massa de 119Da que faz com que o MS/MS esteja no modo NL-119. No entanto, no aminoácido básico citrulina é observado um pico nos 215m/z no modo NL-102 dado que alguns instrumentos produzem uma fragmentação no grupo amónia na fonte de iões do MS/MS [20]. Para o aminoácido glicina, e especificamente para este aminoácido, é necessário o MS/MS operar em NL-56Da para o detectar [51].

As acilcarnitinas partilham em comum uma molécula, que quando submetida à câmara de colisão, produz um ião com massa de 85Da [20, 45]. O MS/MS está na forma de precursor para a análise de 85Da. As acilcarnitinas são categorizadas de grosso modo por acilcarnitinas de cadeia muito longa, longa, média e curta, dependendo do autor.

Durante os dois minutos no qual decorre o processo de análise, o MS/MS é programado de forma a detectar os vários marcadores da seguinte forma [51]:

- acilcarnitinas - 0.00-1.00 minutos sob a forma de *precursor* na massa de 85Da de 215 até 500Da;
- aminoácidos – 1.00-1.45 minutos sob a forma de *neutral loss* 102Da varrendo os 125Da a 300Da;
- aminoácidos – 1.45-1.65 minutos sob a forma de *neutral loss* 119Da varrendo os 150Da a 270Da;
- aminoácidos – 1.65-1.90 minutos sob a forma de *neutral loss* 56Da varrendo os 120Da a 155Da.

Para cada poço da placa, que corresponde a um recém-nascido, obtemos um espectro de massa através do software disponibilizado pelo vendedor do equipamento, que no caso do IGM é o MassLynx. Esse software somente edita o espectro de massa sem possibilidade de rastreio de doenças metabólicas. Este facto fica a dever-se à diversidade de aplicações que têm os MS/MS, sendo o caso do rastreio de doenças metabólicas, uma área de reduzido interesse comercial.

4.2 Caso Português

Em Portugal o programa Nacional de Diagnóstico Precoce teve início em meados de 1979, começando com o rastreio da Fenilcetonúria (PKU), seguido do rastreio do Hipotiroidismo congénito (HC), em 1981 [52].

Entre os anos de 1987 e 1995 foram realizados testes piloto nas doenças da Hiperplasia Congénita das Suprarenais, Deficiência da Biotinidase e Fibrose Quística, os quais foram abandonados devido à não adição de vantagens para a população em geral [53].

Até ao final de 2004 tinham sido rastreados 2 milhões de recém-nascidos, conseguindo detectar perto de mil casos de doenças sendo 723 casos de HC e 214 de PKU. A frequência traduz-se assim de um para, aproximadamente, 3.200 no caso da fenilcetonúria, e de um para, aproximadamente, 11.000 no caso do hipotiroidismo [53]. Actualmente, a taxa de cobertura a nível nacional é de 99.5%, e o tempo médio de início de tratamento é de 12,5 dias, o que representa uma sólida implantação do rastreio a nível nacional [53]. O tratamento destas doenças é possível através de dietas especiais, ou através de tratamento específico.

Em 2003 foram instalados no IGM dois espectrómetros de massa (MS/MS), permitindo a inclusão de um novo conjunto de doenças no Programa de Rastreio Nacional. O IGM definiu a inclusão de 13 doenças genéticas, enumeradas na Tabela 2, que estão geralmente associadas a deficiências de

metabolismo que podem ter consequências graves na saúde dos indivíduos, provocando, nomeadamente, atrasos no desenvolvimento mental.

Fora deste rastreio ficam muitas outras patologias [32], embora o equipamento instalado permita obter informações a diagnosticar. A escolha de um número limitado de doenças prende-se com o facto de existir tratamento ou, no caso de falta de tratamento, existir a possibilidade de melhorar a qualidade de vida dos recém-nascidos.

As doenças incluídas no rastreio pertencem à classe das doenças do ciclo da ureia, da β -oxidação mitocondrial (défice em SCAD, MCAD, LCHAD), aminoacidopatias (PKU, Leucínose, citrulinemia, homocistinúria) e acidúrias orgânicas (propiónica, metilmalónica, 3-hidroxi-3-metilglutárica, glutárica tipo I).

O processo de recolha é efectuado em todo continente e ilhas, recorrendo aos hospitais ou centros de saúde. A recolha deve ser efectuada entre o primeiro e o segundo dia de vida [48], recolhendo várias gotas de sangue do indivíduo através da sua absorção num cartão. Esse cartão tem o nome de Guthrie em honra ao precursor deste sistema. Posteriormente, o cartão é enviado dos centros de saúde ou hospitais para o IGM, no Porto, no qual se procederá ao rastreio. No momento de recolha de sangue do recém-nascido os pais ficam na posse de um número que é parte destacável do cartão Guthrie, que possibilita o acesso em linha ao resultado do rastreio¹. A página acedida dá a possibilidade de inserção de um número que está impresso no destacável, acedendo assim ao resultado do rastreio. No entanto, os pais só têm acesso ao resultado do rastreio caso o recém-nascido não tenha nenhuma das patologias acima indicadas. Caso contrário, o resultado da consulta será “em estudo”, sendo os pais contactados por via telefónica para efectuar uma contra-análise ao recém-nascido.

O IGM garante um acesso restrito e controlado aos dados de forma a preservar a confidencialidade dos cidadãos. Somente quatro pessoas do IGM têm acesso às concentrações obtidas nas análises. Estas pessoas assinaram um termo de confidencialidade junto da Comissão Nacional de Protecção de Dados, devido à informação obtida nos rastreios ser muito delicada, podendo, mesmo, definir o carácter de uma pessoa futuramente. Recentemente, num estudo publicado fazia-se a relação entre um bom e mau trabalhador, através do nível de concentração de uma proteína existente no cérebro dos macacos [54]. Estes resultados são facilmente extrapolados para os humanos [54].

¹ http://www.diagnosticoprecoce.org/result_entry.php

4.3 Lacunas/problemas

A tecnologia MS/MS permite detectar uma grande variedade de metabolitos a partir de uma mesma amostra de sangue. A realização deste tipo de rastreio tem o inconveniente de produzir uma quantidade enorme de informação: cerca de 60 concentrações de diferentes marcadores por recém-nascido, como se pode ver na Figura 21. Cada linha da figura corresponde a um indivíduo, estando presente nas primeiras quatro linhas as concentrações relativas aos controlos CDC.

Para dificultar ainda mais o processo de análise, cada concentração terá que estar compreendida dentro de um determinado limite, considerado normal para a idade e para o peso. No entanto, existe alguma dificuldade em definir esses limites, fora dos quais se decide que determinado recém-nascido é possuidor ou não, de uma ou mais doenças. Definindo limites muito pequenos, estamos a aumentar o número de falsos positivos, mas se procedermos de forma inversa e dilatarmos muito o limite, corremos o risco de obter falsos negativos.

É também necessário monitorizar a evolução das concentrações dos marcadores ao longo do tempo, para testar se não existe nenhum desvio gradual ou em forma de pico em relação à média diária. Estes desvios indicam que o equipamento tem alguma falha de calibração ou execução.

Considerando que nascem cerca de 600 crianças por dia em Portugal [52] torna-se quase impossível realizar o rastreio sem a ajuda de uma aplicação informática que trate e organize a informação.

Actualmente existem algumas aplicações que ajudam o operador no rastreio, mas nenhuma preenche as recomendações do CDC apresentadas anteriormente.

A aplicação Neogram da Perkin [55], mostra ser um produto com muita qualidade, contendo muitas funcionalidades recomendadas pelo CDC. Permite calcular os limites dos marcadores de forma dinâmica, através do cálculo dos percentis sobre os marcadores passados devidamente validados pelo investigador. Esta forma de cálculo é sugerida pelo CDC, podendo, também, definir esses limites de forma empírica. Os limites empíricos podem ser obtidos nos relatórios anuais publicados pelo CDC [50] obtidos através das análises efectuadas às concentrações enviadas pelos laboratórios ao longo do ano. Contudo, o Neogram, não contém o módulo de testar a evolução das concentrações médias ao longo do tempo.

File Name	Sample Nam	Asp conc	Glu conc	Gly conc	Ala conc	Cit conc	Met conc	X-Leu conc	Tyr conc	Phe conc	Orn conc	Arg conc
of 'Unknown'		6	180	290	164	6	6	72	61	20	25	4
of 'Unknown'		157	736	1400	819	29	50	450	400	130	215	32
03.11.26A(1)	CDC 461	31,50	452,87	1041,69	553,35	26,11	25,44	257,67	80,74	110,08	142,99	12,33
03.11.26A(1)	CDC 462	34,80	475,50	1147,75	717,77	28,59	35,02	294,88	78,18	122,52	170,00	14,74
03.11.26A(1)	CDC 463	64,22	531,03	1075,35	549,40	30,55	28,77	250,61	112,38	111,60	147,40	19,23
03.11.26A(1)	CDC 464	55,10	540,69	1172,30	659,08	30,56	35,77	319,56	120,35	135,56	169,69	18,34
03.11.26A(1)	2326103.00	47,83	357,93	764,39	423,21	14,20	12,04	167,22	150,16	51,38	105,27	12,31
03.11.26A(1)	2327002.00	62,08	372,58	654,84	310,73	10,36	13,66	142,37	134,19	51,06	96,10	9,33
03.11.26A(1)	2327003.00	67,98	445,51	676,43	308,85	14,52	12,31	146,42	141,52	46,30	88,59	8,04
03.11.26A(1)	2327004.00	43,48	351,67	574,87	309,37	10,38	11,88	168,73	212,50	53,68	75,12	10,94
03.11.26A(1)	2327005.00	57,81	435,62	783,81	476,29	10,91	18,06	122,93	105,64	61,66	102,97	14,15
03.11.26A(1)	2327006.00	41,45	462,28	644,90	492,05	13,77	22,00	208,22	108,09	43,98	78,13	13,60
03.11.26A(1)	2327007.00	51,32	459,21	588,29	361,54	14,45	13,74	187,67	152,52	56,04	80,12	9,95
03.11.26A(1)	2327008.00	32,76	349,00	316,57	352,48	12,01	19,99	156,67	111,54	38,02	50,27	8,10
03.11.26A(1)	2327009.00	21,19	399,68	423,40	278,35	7,45	10,65	141,84	113,96	58,86	41,12	5,89
03.11.26A(1)	2327010.00	90,65	501,18	1004,29	733,89	16,26	23,40	235,23	136,79	47,89	143,21	13,04
03.11.26A(1)	2327011.00	64,26	580,92	1219,05	672,71	13,79	32,78	326,96	267,08	113,67	161,78	14,88
03.11.26A(1)	2327012.00	41,93	343,07	794,71	487,38	10,10	17,49	211,36	272,79	58,45	79,05	8,05
03.11.26A(1)	2327013.00	47,23	300,55	519,21	414,05	13,49	17,48	199,74	125,90	45,40	93,57	11,97
03.11.26A(1)	2327014.00	39,73	417,76	701,19	342,85	9,53	21,92	158,67	99,33	51,23	66,42	3,05
03.11.26A(1)	2327015.00	44,56	341,65	430,37	267,31	15,53	15,14	147,43	103,77	42,88	71,14	9,38
03.11.26A(1)	2327016.00	35,76	303,58	464,94	340,17	9,03	16,18	213,95	132,42	55,68	54,49	7,16
03.11.26A(1)	2327017.00	40,07	346,68	686,34	546,35	12,63	34,54	217,99	121,11	48,43	102,58	8,05
03.11.26A(1)	2327018.00	112,33	527,00	1104,95	632,29	15,14	22,50	224,36	225,44	69,91	213,42	15,48
03.11.26A(1)	2327019.00	46,93	394,10	751,46	365,32	10,44	23,46	184,99	122,57	53,56	98,16	10,25
03.11.26A(1)	2327020.00	57,53	405,78	694,51	546,41	17,89	24,21	222,54	158,92	64,50	109,87	12,78
03.11.26A(1)	2327021.00	56,68	441,37	590,88	294,80	11,55	20,44	155,14	115,94	51,92	83,81	7,86
03.11.26A(1)	2327022.00	46,45	363,15	582,33	459,47	17,18	24,10	188,71	108,43	48,24	99,58	14,72
03.11.26A(1)	2327023.00	65,20	349,35	709,55	497,62	14,55	14,20	226,80	134,86	62,35	120,46	13,10
03.11.26A(1)	2327024.00	59,74	325,14	671,88	486,99	22,53	15,93	192,30	196,09	45,31	124,47	21,66
03.11.26A(1)	2327026.00	52,20	438,54	782,49	452,21	26,09	18,88	156,53	84,11	51,80	94,54	10,93
03.11.26A(1)	2327027.00	64,99	357,67	964,39	421,99	29,50	22,17	229,47	149,54	70,24	122,96	12,97
03.11.26A(1)	2327028.00	18,95	214,34	506,20	257,84	11,34	17,09	148,25	67,13	48,83	48,20	9,37
03.11.26A(1)	2327029.00	31,20	271,60	581,91	206,69	10,99	23,02	155,75	106,26	42,35	70,97	9,35
03.11.26A(1)	2327030.00	70,45	380,12	806,61	534,63	71,89	24,69	227,52	133,05	78,85	105,41	10,77
03.11.26A(1)	2327031.00	19,21	252,88	474,40	283,34	8,08	19,31	112,15	98,03	42,67	41,02	4,72
03.11.26A(1)	2327032.00	119,84	569,54	736,47	483,61	15,38	22,21	233,48	122,64	66,80	132,47	12,67
03.11.26A(1)	2327033.00	33,95	341,71	538,58	395,85	10,12	38,89	154,77	108,95	57,04	58,80	6,56
03.11.26A(1)	2327034.00	31,59	405,72	545,79	347,65	11,21	16,10	146,40	109,82	45,73	45,33	6,31
03.11.26A(1)	2327035.00	55,09	459,35	751,05	489,86	12,82	34,84	228,63	125,72	55,14	97,91	9,40
03.11.26A(1)	2327036.00	39,39	318,46	640,10	453,48	13,29	17,09	182,14	146,65	50,95	85,72	12,10
03.11.26A(1)	2327037.00	34,47	243,61	484,31	435,65	12,22	29,23	178,18	125,06	53,68	70,38	9,80
03.11.26A(1)	2327038.00	52,83	493,19	749,00	404,51	10,54	20,02	190,32	64,68	62,80	95,35	8,67
03.11.26A(1)	2327039.00	27,81	243,89	286,36	258,44	6,87	12,71	129,78	78,30	27,86	47,24	3,63
03.11.26A(1)	2327040.00	29,15	350,56	504,56	470,60	10,09	25,74	290,29	432,57	67,58	86,76	13,27
03.11.26A(1)	2327041.00	33,74	312,27	404,54	301,58	7,52	13,00	133,79	174,41	37,64	48,30	7,91
03.11.26A(1)	2327042.00	28,24	237,04	365,18	253,87	11,77	18,10	136,44	90,49	40,99	56,99	10,11
03.11.26A(1)	2327043.00	20,81	267,84	468,43	304,69	10,36	17,78	137,44	88,93	47,09	49,25	4,37
03.11.26A(1)	2327044.00	92,37	433,55	1105,38	508,96	9,30	17,86	153,53	147,27	56,55	194,71	18,67
03.11.26A(1)	2327045.00	71,76	320,68	744,06	506,17	18,54	23,81	192,98	112,97	56,58	122,91	13,13
03.11.26A(1)	2327046.00	74,72	465,18	955,83	533,51	12,27	33,88	215,82	138,60	76,12	132,43	18,78
03.11.26A(1)	2327047.00	26,06	356,75	519,55	368,00	7,93	10,70	144,62	98,23	32,60	48,32	6,36

Figura 21 – Concentrações obtidas na análise realizada a 40 indivíduos presentes numa placa através do MS/MS. Estes dados são gerados pela aplicação ChemoView. A tabela foi cortada por motivos de espaço. São visualizados apenas onze marcadores dos sessenta actualmente medidos.

No sistema CAMPA [51], os limites foram calculados através da análise dos marcadores de 1100 recém-nascidos, obtendo-se assim um valor empírico. Os valores obtidos são depois confrontados com os marcadores em análise, obtendo-se os indivíduos que têm valores fora dos limites. Esta análise, não estando errada, vai contra a recomendação do CDC que sugere que os limites sejam obtidos através dos percentis ao longo do tempo [22].

Um outro sistema desenvolvido por Baumgartner [56], optou por definir os limites através de redes neuronais. Foram obtidos bons resultados de forma a conseguirem baixas percentagens de falsos positivos e falsos negativos. No entanto, este sistema de redes neuronais só foi aplicado nas doenças PKU e MCADD, devido à existência de uma quantidade razoável de casos detectados.

Para as outras doenças, não existem casos detectados em número suficiente para “ensinar” as redes neuronais. Este facto fica-se a dever ao estudo estar centralizado em doenças raras, dificultando ainda mais o trabalho do investigador.

O IGM tem instalado de uma aplicação, ChemoView, que permite gerar ficheiros em formato Excel a partir dos espectros de massa obtidos no MS/MS através da aplicação Masslynx. Esta aplicação permite, somente, definir os limites de forma empírica, assinalando as concentrações fora do limite a vermelho. O investigador terá de analisar toda a informação gerada de forma manual recorrendo somente aos ficheiros Excel, tornando-se muito cansativo e susceptível a erros. O IGM não tem qualquer tipo de aplicação informática que auxilie no rastreio dos recém-nascidos.

Mediante esta análise foi proposto o desenvolvimento de um sistema de informação, suportado por ferramentas matemáticas e estatísticas, que automatiza o diagnóstico dos recém-nascidos, facilitando o rastreio nos laboratórios (Figura 22).

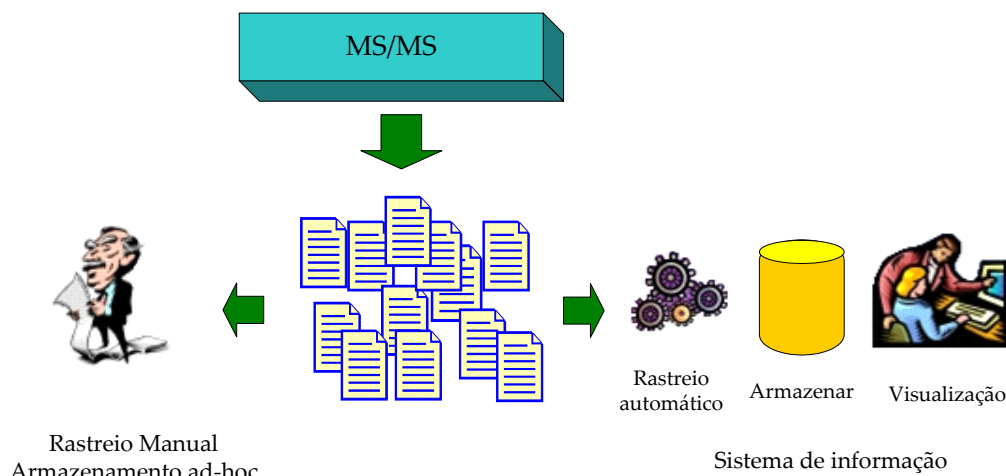


Figura 22 – Duas formas de rastreio possível.

As ferramentas estatísticas incorporadas no sistema, permitem uma actualização automática e diária dos limites dos marcadores, através do cálculo dos seus percentis, relativos a valores passados e validados.

4.4 Sumário

A realização do rastreio de doenças em recém-nascidos é efectua em Portugal desde a década de 70. No entanto, só era possível rastrear duas doenças devido à tecnologia existente no IGM, ou devido ao benefício custo/vantagem para a população.

Recentemente, com a aquisição de dois espectrómetros de massa em tandem é possível rastrear um elevado conjunto de doenças. Devido à grande quantidade de informação gerada pelos meios de análise é necessário recorrer a sistemas de informação para auxiliar o investigador no diagnóstico. No entanto, as ferramentas disponíveis não preenchem os requisitos de controlo impostos aos rastreios. Para colmatar o problema é proposto um sistema de informação que efectua a leitura, processamento, controlo e análise dos dados obtidos através do MS/MS.

5 NeoScreen - ferramenta de auxílio ao diagnóstico de doenças metabólicas em recém-nascidos.

Na análise dos processos geradores de uma grande quantidade de informação, é imprescindível desenvolver aplicações que ajudem no suporte à decisão e auxiliem em muito, um trabalho que é extremamente monótono e muito sujeito a erros.

Para tal, desenvolveu-se uma aplicação que tem por nome NeoScreen e que executa rastreios de forma automática, eliminando grande parte das tarefas monótonas e sujeitas a erros, por parte dos investigadores, simplificando tarefas e ajudando no suporte à decisão. Tem como principal objectivo a identificação de potenciais casos de doenças metabólicas em recém-nascidos.

No presente capítulo explica-se os requisitos que foram impostos ao desenvolvimento da aplicação, assim como uma breve explicação da sua arquitectura, implementação e descrição da base de dados.

5.1 Requisitos

A elaboração dos requisitos para a aplicação foram traçados pela equipa de investigadores do IGM, cuja experiência no rastreio e detecção de doenças metabólicas, foi fundamental para definir as funcionalidades que uma determinada aplicação necessita para proceder a um correcto rastreio.

Para simplificar o planeamento da aplicação recorreu-se a um diagrama de casos de utilização (Figura 23), em notação *Unified Modeling Language* (UML) [57], que define os requisitos básicos do sistema.

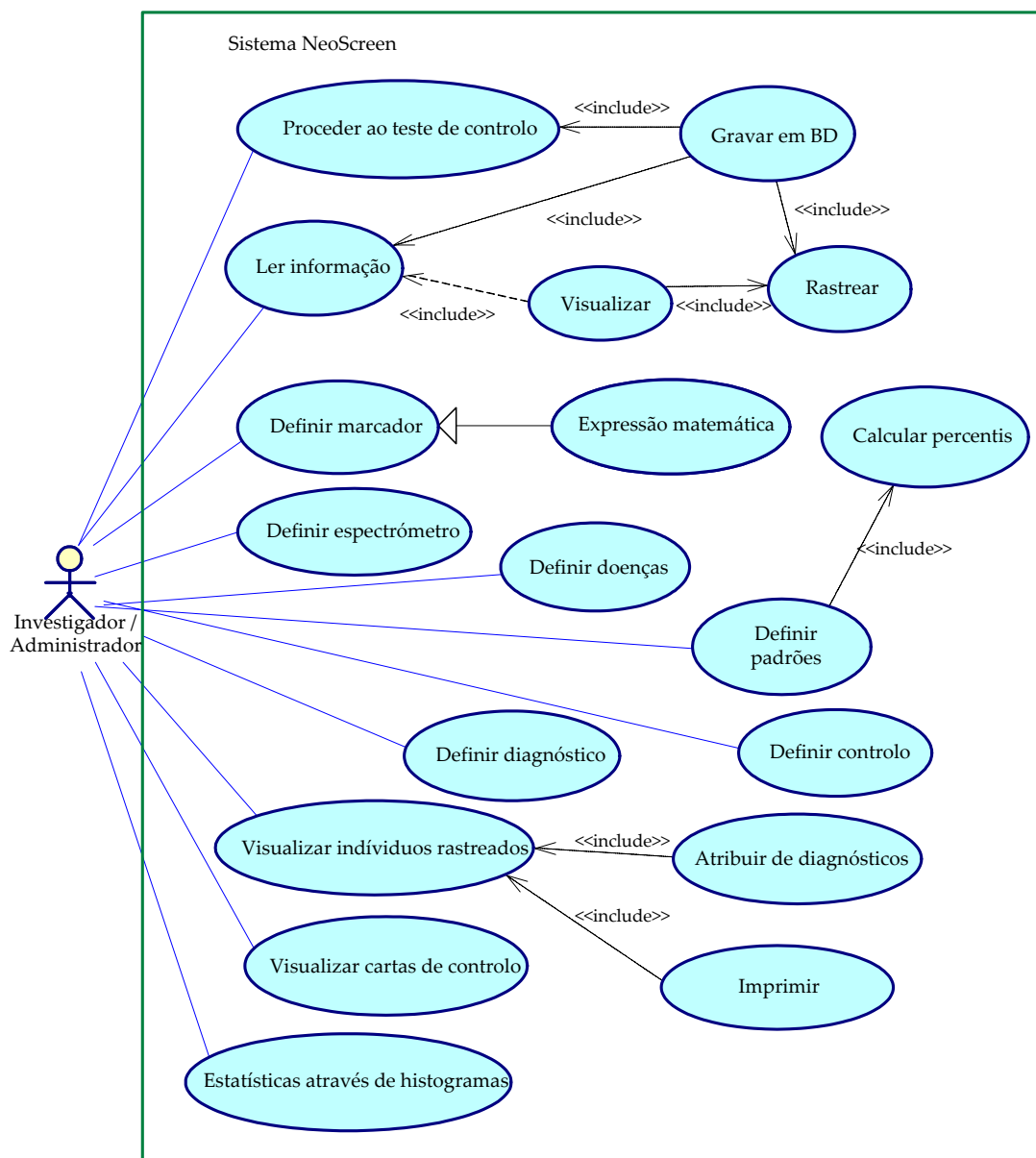


Figura 23 – Diagrama de casos de utilização para o sistema NeoScreen.

5.1.1 Descrição dos actores

O investigador/administrador é o actor principal tendo poderes para configurar todo o sistema. A segurança é realizada através da base de dados, mediante a introdução de uma palavra-chave para estabelecer a ligação.

A decisão de nomear todos os utilizadores como administradores, deveu-se ao facto de aplicação possuir um número muito restrito de utilizadores. Todos os utilizadores da aplicação estão devidamente registados na base de dados, sendo responsáveis pela sua gestão.

5.1.2 Descrição dos casos de utilização

A definição de marcadores permite ao utilizador a possibilidade de os criar, de os alterar ou de os eliminar. Os marcadores contêm um conjunto de propriedades, como a forma de impor limites, qual a unidade ou se faz parte de um grupo de marcadores de intensidade.

A possibilidade de criar expressões matemáticas com os marcadores terá de ser contemplada, possibilitando assim a criação de rácios ou outras operações entre diferentes marcadores.

A criação de unidades permite ao utilizador definir unidades por extenso para posteriormente associá-las aos marcadores.

Com a definição de espectrómetros o utilizador poderá tratar estatisticamente a informação por espectrómetro ou no seu global. Com esta opção, o utilizador tem a possibilidade de observar se existe alguma discrepância ao nível das medições médias entre espectrómetros.

A definição de doenças permite ao utilizador a possibilidade de criar, alterar ou eliminar doenças possíveis de rastrear no sistema. A cada doença está associado um ou mais marcadores, e para cada marcador, tem de se definir quais os limites que não podem ser violados de forma a activar a doença. Pode-se definir se é um limite inferior, superior ou ambos.

A definição dos padrões vai permitir ao utilizador criar, alterar ou eliminar padrões. É com estes padrões que o utilizador confronta a informação lida nas tabelas originais de forma a realizar os rastreios. Cada padrão definido pelo utilizador contém a totalidade dos marcadores, assim como a forma de definir os limites, se de forma empírica ou calculada por percentis. Existem dois níveis de severidade para o limite inferior (Muito Baixo e Baixo) e dois para limite superior (Alto e Muito Alto), como se pode ver na Figura 24.

No caso dos limites empíricos, o nível 1 no limite inferior e superior é definido pelo utilizador no próprio marcador, enquanto o nível 2 é determinado por percentagem. No caso dos percentis, os níveis são definidos pela introdução dos valores dos percentis nos próprios marcadores. Os cálculos dos limites ficam a cargo da aplicação, sendo recalculados todos os dias, tornando-se assim um sistema dinâmico. O sistema terá de permitir a definição dos indivíduos que servem de fonte para o cálculo dos percentis.

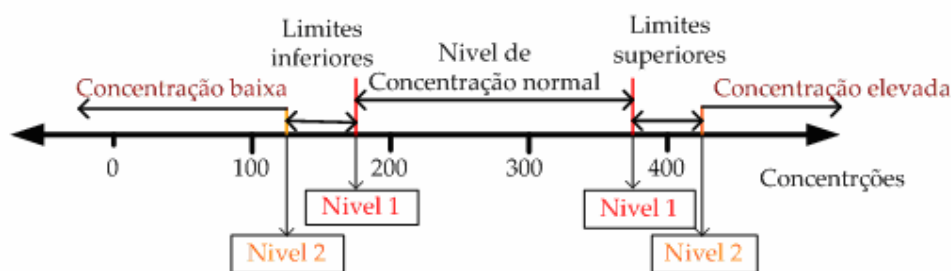


Figura 24 – Conceito de limite que se aplica aos marcadores responsáveis pela suspeita de doenças nos recém-nascidos.

O diagrama de caso “Calcular percentis” permite à aplicação obter os valores dos percentis que foram definidos/alterados nas regras, ou obter os valores através do obrigatório cálculo diário.

Na definição de controlos CDC o utilizador cria, elimina ou altera os controlos. A cada controlo está associado um ou mais marcadores, os quais são distribuídos e definidos pelo CDC. Estes controlos terão de estar acima de um determinado limite definido pelo utilizador ou calculado através do percentil.

Na visualização dos indivíduos rastreados, o utilizador tem a possibilidade de consultar os indivíduos por doença, dia, rastreio ou através da concentração de determinado marcador. Os indivíduos são agrupados por sete categorias, sendo elas:

- Muito suspeitos – Todos os indivíduos suspeitos de uma ou mais doenças. Neste caso, pelo menos um marcador que pertence a uma doença diagnosticada, tem uma concentração que viola o nível dois;
- Suspeitos – Todos os indivíduos suspeitos de uma ou mais doenças, mas que as concentrações dos marcadores referenciados nas doenças estão compreendidas entre o nível um e o nível dois;
- Não suspeitos – Todos os indivíduos que têm as concentrações dos marcadores normais, compreendidas no nível um;
- Valores muito altos mas sem doenças – Todos os indivíduos que têm concentrações acima do nível dois do limite superior, mas os marcadores em causa não estão associados às doenças;
- Valores altos mas sem doenças – Todos os indivíduos que têm concentrações que estão compreendidas entre o nível um e o nível dois do limite superior, mas esses marcadores não estão associados às doenças;

- Valores baixos mas sem doenças – Todos os indivíduos com concentrações inferiores ao nível um e ao nível dois do limite inferior, mas os marcadores não estão associados às doenças;
- Falha de intensidade – Todos os indivíduos que contêm as concentrações dos marcadores intensidade abaixo do limite imposto nas regras. Para os indivíduos em causa terá de se proceder a um novo rastreio;
- Controlo CDC – contém os controlos CDC analisados para o dia em consulta.

O utilizador ao seleccionar determinado indivíduo obtém variada informação relativa às concentrações referentes aos marcadores, possíveis doenças, limites impostos aos marcadores, entre outros dados possíveis.

A atribuição de diagnósticos permite ao utilizador definir diagnósticos aos indivíduos em visualização. Permite também confirmar doenças suspeitas e definir se determinado indivíduo vai ser incluído no lote de indivíduos que servem como referência para o cálculo dos percentis. Torna-se necessário que os indivíduos envolvidos no cálculo dos limites dos percentis tenham as concentrações dentro dos valores normais, caso contrário, corre-se o risco de introduzir erros no cálculo dos limites.

A impressão permite ao utilizador escolher os dados referentes ao indivíduo que deseja imprimir, como a sua identificação, marcadores e suas concentrações, doenças suspeitas ou confirmadas e possíveis observações.

Na visualização das cartas de controlo, o utilizador poderá observar a evolução das médias diárias das concentrações dos marcadores, através de um histograma que evolui diariamente. Existe um histograma para cada marcador. Os valores que servem de fonte para o cálculo das médias são obtidos através de uma ou mais categorias seleccionadas pelo utilizador. O utilizador poderá seleccionar que só os indivíduos pertencentes às categorias normais e valores altos mas sem doenças sejam incluídos nos cálculos das médias, por exemplo.

Nas estatísticas através de histogramas o utilizador poderá obter variada informação, como o número de indivíduos suspeitos ou confirmados de determinada doença, analisados por determinado espectrómetro, pertencentes a uma determinada categoria ou possuidores de determinado diagnóstico. Os histogramas podem ser consultados por dias, meses ou anos.

No diagrama de caso “Ler informação” o utilizador pode ler os ficheiros devidamente formatos e obtidos através da aplicação *ChemoView*, para o NeoScreen. O utilizador tem a possibilidade de escolher a regra que vai confrontar com os indivíduos lidos do ficheiro. A aplicação, no momento da leitura, constrói várias listas, identificando quais os marcadores reconhecidos e

rejeitados, quais os controlos CDC identificados e rejeitados, quais os indivíduos que falham nos marcadores de intensidade e os indivíduos aceites para diagnóstico.

O utilizador pode gravar a informação na base de dados do sistema ou só proceder à sua análise. No diagrama de caso “Gravar em BD” o utilizador grava os indivíduos e seus marcadores na base de dados para posterior consulta e análise. Terá de se executar primeiro o rastreio de forma a diagnosticar possíveis doenças, atribuir categoria e diagnóstico. O rastreio é executado confrontando a regra escolhida no momento da leitura com as concentrações presentes nos indivíduos em análise.

O diagrama de caso “Proceder ao teste de controlo” é despoletado pelo sistema aquando da gravação da informação em base de dados. O teste consiste em comparar a média diária do marcador com o desvio padrão obtido ao longo do tempo para esse mesmo marcador. Os testes a efectuar são:

- O valor individual médio do dia está acima ou abaixo três vezes em relação ao desvio padrão global;
- O valor médio de dois dias seguidos está acima ou abaixo duas vezes em relação ao desvio padrão;
- O valor individual médio de cinco dias consecutivos está acima ou abaixo do valor da média global para determinado marcador;
- O valor individual médio de cinco dias consecutivos tem subido progressivamente ou descido gradualmente em relação à média global do marcador.

No diagrama de caso “visualizar” o utilizador emprega somente as potencialidades da aplicação para visualizar os indivíduos lidos através dos ficheiros gerados pelo *ChemoView*, necessitando ainda assim de executar o rastreio para diagnosticar possíveis doenças. Não é gravada qualquer informação em base de dados, com este tipo de diagrama de caso.

5.2 Base de dados

Para projectar a base de dados necessária ao suporte da aplicação desenhamos um diagrama conceptual, representado na Figura 25.

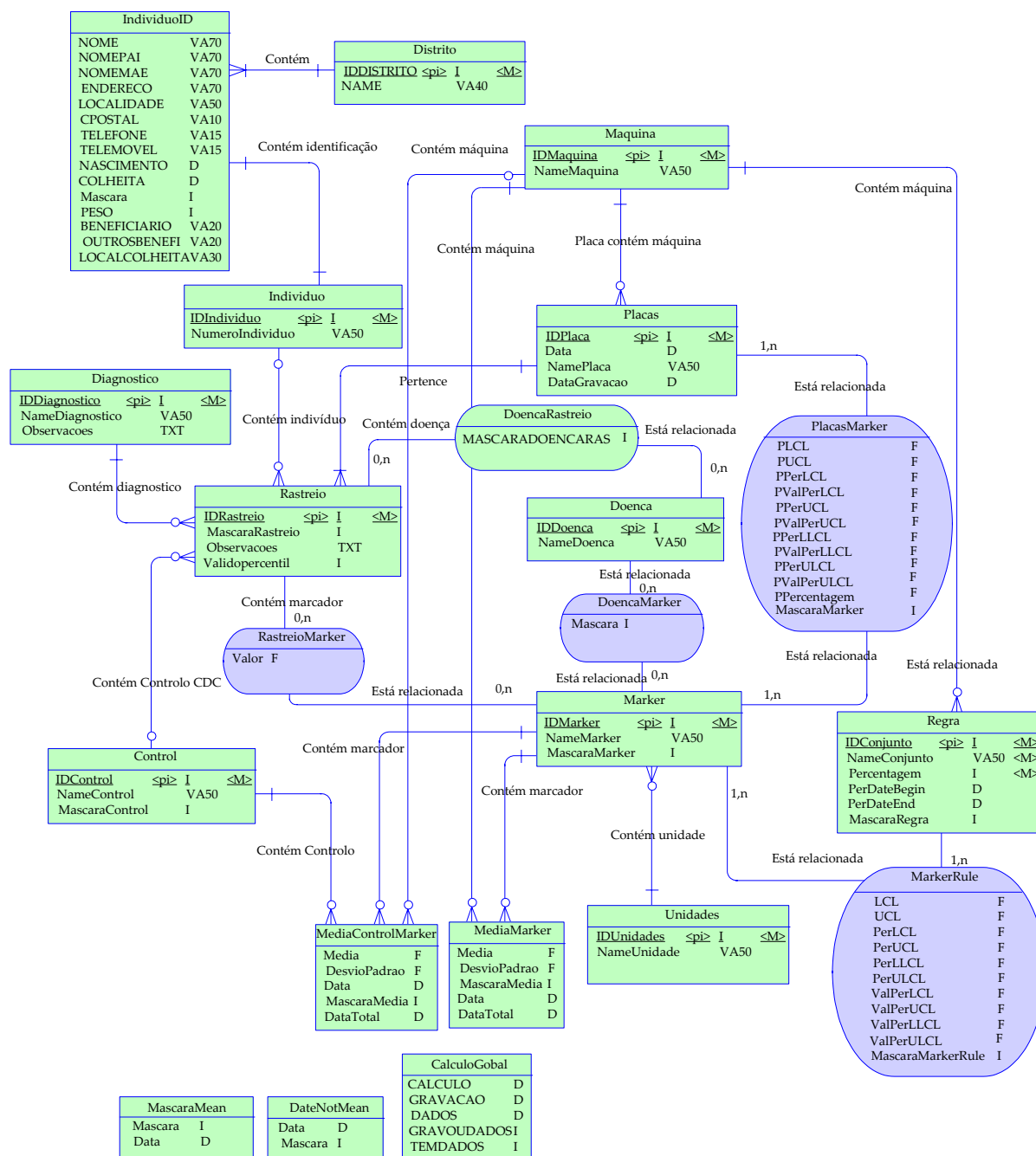


Figura 25 – Diagrama conceptual da base de dados.

O diagrama conceptual é constituído por um conjunto de entidades e relacionamentos. A explicação das entidades e sua utilização está descrita na Tabela 4.

Tabela 4 – Descrição das entidades

Entidade	Descrição	Ocorrência
Rastreio	A entidade que define a presença no sistema de determinado indivíduo ou controlo CDC	Existe um rastreio por cada indivíduo ou controlo CDC admitido no sistema.
Placas	O conjunto de indivíduos e/ou controlo CDC que estavam presentes na mesma placa no momento da análise no MS/MS.	No momento da importação de um ficheiro Excel para a análise no NeoScreen, esse ficheiro fica rotulado com pertencente a uma determinada placa com identificação única.
Individuo	Identificação única e simples de um indivíduo.	Cada indivíduo está identificado por um número que será obtido no momento da importação do ficheiro.
Diagnostico	Propriedade atribuída a cada rastreio executado pelo sistema.	No momento de leitura será atribuída um diagnóstico padrão, podendo o utilizador à posteriori alterar o diagnóstico mediante a realização de uma análise.
Maquina	Identificação para os MS/MS existentes em laboratório	
Control	Controlos CDC disponíveis no laboratório e utilizados nas análises.	Cada ficheiro importado para a aplicação contém em média quatro controlos CDC.
Doenca	Nome das doenças rastreadas pela aplicação.	
Marker	Nome dos marcadores reconhecidos pela aplicação ou marcadores obtidos através de expressões matemáticas	
Unidades	Qual a unidade que tem cada marcador	
Regra	Conjunto de marcadores onde estão definidos os limites e outras propriedades necessárias para efectuar o rastreio dos indivíduos.	
MediaControlMarker	Contém quais os marcadores presentes para cada controlo CDC. As médias diárias estão incluídas também nesta entidade assim como a média global e desvio padrão por marcador.	No momento da importação dos ficheiros as médias dos marcadores que correspondem aos controlos CDC são calculados. As médias globais e desvios padrões são recalculados.
MediaMarker	Contém as médias diárias por marcador assim como a média global e desvio padrão por marcador.	No momento da importação dos ficheiros as médias dos marcadores são calculados e as médias globais e desvios padrões são recalculados.

Além das entidades, podemos sugerir também as seguintes entidades fracas descritas na Tabela 5.

Tabela 5 – Descrição das entidades fracas

Entidade	Descrição	Ocorrência
IndividuoID	Identificação pessoal do indivíduo.	Quando o utilizador necessita dos seus dados por motivos de impressão.
Distrito	Descrição de distritos.	
CalculoGlobal	Informação necessária para o calculo dos percentis.	Todos os dias o sistema verifica se já ocorreu o cálculo dos percentis para o dia em questão.
MascaraMean	Categorias escolhidas para o calculo das médias diárias e globais dos marcadores.	
DateNotMean	Informação necessária para o calculo das médias diárias dos marcadores	Todos os dias o sistema verifica se já ocorreu o calculo das médias diárias e globais para os marcadores presentes no sistema

A descrição dos relacionamentos com atributos é apresentada na Tabela 6.

Tabela 6 - Descrição dos relacionamentos com atributos

Relacionamentos	Descrição	Ocorrência
DoencaRastreio	Contém as doenças que estão associados a determinado rastreio.	Quando um individuo rastreado é suspeito de possuir alguma doença definida no sistema.
RastreioMarker	Contém os marcadores existentes e as suas concentrações para cada rastreio.	
DoencaMarker	Contém quais os marcadores que estão relacionados com as doenças.	
PlacasMarker	Contém quais os limites em percentil ou empíricos usados no rastreio de determinada placa.	

A descrição dos relacionamentos sem atributos é apresentada na Tabela 7

Tabela 7 – Descrição dos relacionamentos sem atributos

Entidade	Multiplicidade	Relacionamento	Entidade	Multiplicidade
Rastreio	0..1	Contém indivíduo	Individuo	0..M
	1..1	Contém diagnóstico	Diagnóstico	0..M
	0..1	Contém controlo CDC	Control	0..M
	0..N	Contém marcador	Marker	0..M
	1..1	Pertence	Placas	1..M
	0..N	Contém doença	Doenca	0..M
Control	0..N	Contém controlo CDC	Rastreio	0..1
	0..N	Contém Controlo	MediaControlMarker	1..1
Individuo	0..N	Contém indivíduo	Rastreio	0..1
	1..1	Contém identificação	IndividuoID	1..1
Maquina	0..N	Contém máquina	MediaControlMarker	0..1
	0..N	Contém máquina	MediaMarker	1..1
	0..N	Placa contém máquina	Placas	1..1
Placas	0..N	Contém máquina	Regra	1..1
	1..1	Placa contém máquina	Maquina	0..M
	1..N	PlacasMarker	Marker	1..M
	1..N	Pertence	Rastreio	1..1
Doença	0..N	Está relacionado	Rastreio	0..M
	0..N	Está relacionado	Marker	0..M
Maker	0..N	Está relacionado	Rastreio	0..M
	0..N	Contém marcador	MediaControlMarker	1..1
	0..N	Contém marcador	MediaMarker	1..1
	1..1	Contém unidade	Unidades	0..M
	1..N	Está relacionada	Regra	1..M
	1..N	Está relacionado	Placas	1..M
Regra	1..1	Contém máquina	Maquina	0..M
	1..N	Está relacionada	Marker	1..M

Um dos aspectos mais importantes a ter em conta em aplicações que envolvam base de dados é o seu crescimento. Mediante o crescimento é que se dimensiona o espaço necessário para o bom funcionamento da aplicação. Os cálculos pormenorizados para o crescimento da base de dados do NeoScreen

encontram-se no Anexo 2. O tamanho inicial será dado pelo total da Tabela 12 e a taxa de crescimento pelo total da Tabela 13. Os cálculos foram obtidos supondo que são introduzidos valores referentes a 600 recém-nascidos por dia.

Com os resultados fornecidos pelas tabelas no anexo 3 temos os seguintes valores:

- Tamanho inicial: 86,022 KBytes;
- Crescimento por cada mês: 32,9 MBytes.

A escolha do motor de base de dados é sempre uma tarefa que exige muita ponderação. Existem vários à escolha, mas é necessário e de extrema importância, escolher um que nos ofereça uma boa robustez e nos transmita confiança. Podemos dividir os motores de bases de dados em duas categorias, *open source* e proprietárias. No caso das *open source* temos o MySQL [58] e a PostgreSQL [59] sendo as duas mais conhecidas. As proprietárias destacam-se a Microsoft SQL Server, Oracle e DB2 da IBM.

Existe uma excepção quanto ao motor de base de dados Microsoft. A Microsoft disponibiliza uma versão gratuita, impondo somente algumas limitações de utilização. O motor de base de dados em causa é a *Microsoft SQL Server 2000 Desktop Engine* (MSDE 2000) [60] que está disponível gratuitamente. No entanto não suporta mais de 25 ligações em simultâneo podendo conter 16 base de dados não superiores a 2GB. Não contém nenhuma aplicação de interacção com o motor de base de dados, necessitando recorrer a outras aplicações para interagir com o motor de base de dados. Pode-se, no entanto, desenvolver a nossa própria ferramenta para o efeito. Este motor de base de dados possui o mesmo núcleo do SQL Server 2000 da Microsoft, dispondo também do mesmo suporte técnico.

A PostgreSQL é um motor de base de dados que à data da criação do NeoScreen necessitava de um emulador de Unix para poder correr em sistemas operativos Windows. O lançamento da versão 8.0, em Janeiro de 2005, vem colmatar esse inconveniente podendo neste momento correr em sistemas operativos Windows sem necessitar de qualquer tipo de emulador [61]. É de referir que a entidade promotora do PostgreSQL não impõe qualquer limitação em termos de licença, sendo uma solução complementa gratuita, mesmo que o PostgreSQL seja utilizado num produto com fins lucrativos.

O motor de base de dados MySQL, o qual é abrangido pela licença *General Public License* (GPL), só é gratuito se disponibilizarmos o código da aplicação desenvolvida em conjunto com o MySQL. Não contém alguns dos requisitos necessários à realização da aplicação NeoScreen como por exemplo as *stored procedures* ou os *triggers*, limitando logo à partida a utilização do MySQL [62].

No entanto, também não havia interesse por parte do IEETA em disponibilizar o código da aplicação NeoScreen para ficar abrangido pelo GPL.

Quanto aos motores de bases de dados Oracle, DB2 e SQL Server, encontram-se os três na mesma gama, dependendo agora do tipo de necessidades, assistência ou verba disponível para o efeito.

Para uma aplicação como a NeoScreen, a qual vai tratar informação muito sensível de ponto de vista ético, temos que dar um peso muito significativo à segurança e fiabilidade, pelo que a nossa escolha foi para um sistema proprietário. Perante estes argumentos, a escolha foi para o MSDE devido a ser uma solução gratuita, robusta e dispendo do suporte técnico da Microsoft. O MSDE permite-nos ter a aplicação NeoScreen em funcionamento pelo menos quatro anos sem necessidade de recorrer a qualquer licença, sendo a limitação de utilizadores não limitativa, devido a ser praticamente impossível um laboratório ter um conjunto de 25 investigadores a trabalhar sobre a base de dados.

Houve, no entanto, a necessidade de desenvolver uma aplicação de comunicação com o motor de base de dados. Esta ferramenta de assistência foi desenvolvida também no âmbito do projecto (Figura 26) para fornecer um conjunto de serviços ao utilizador. Nos serviços incluem-se as cópias de segurança, definições de usuários da base de dados e geradores de relatórios das tabelas pertencentes a determinada base de dados.

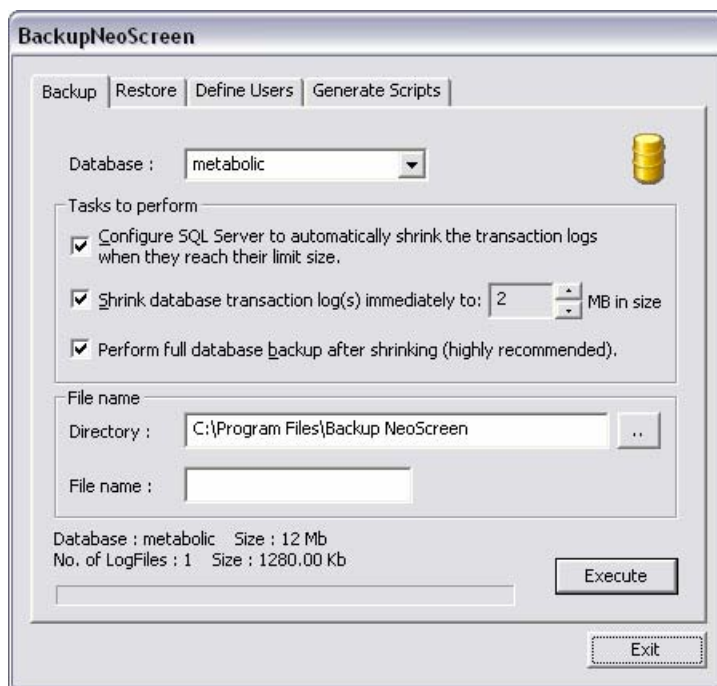


Figura 26 - Ferramenta de assistência ao motor de base de dados.

Para facilitar a tarefa de desenvolvimento, a Microsoft disponibiliza um conjunto de objectos que têm por nome *Distributed Management Objects* (DMO) distribuídos através de uma DLL, a *sqldmo.dll*. Este conjunto de objectos permite ao programador proceder a um conjunto de tarefas, como por exemplo, gerar *scripts*, efectuar cópias de segurança, criar/alterar/eliminar contas de utilizador, entre outras coisas, com relativa facilidade.

5.3 Implementação

Na definição da arquitectura da aplicação NeoScreen teve-se o cuidado de tentar separar as várias funcionalidades por blocos. Para tal, foram definidos quatro blocos principais responsáveis por todas as ferramentas disponíveis no NeoScreen, como mostra a Figura 27.

Os blocos definidos são:

- Aquisição - responsável pela leitura dos ficheiros para o NeoScreen;
- Processamento - responsável pelo rastreio dos indivíduos provenientes da Aquisição, assim como a sua gravação em base de dados;
- Visualização - responsável pela visualização da informação, assim como algumas caixas de diálogo também têm esta função;
- Configuração - engloba todas as possibilidades de definição de parâmetros.

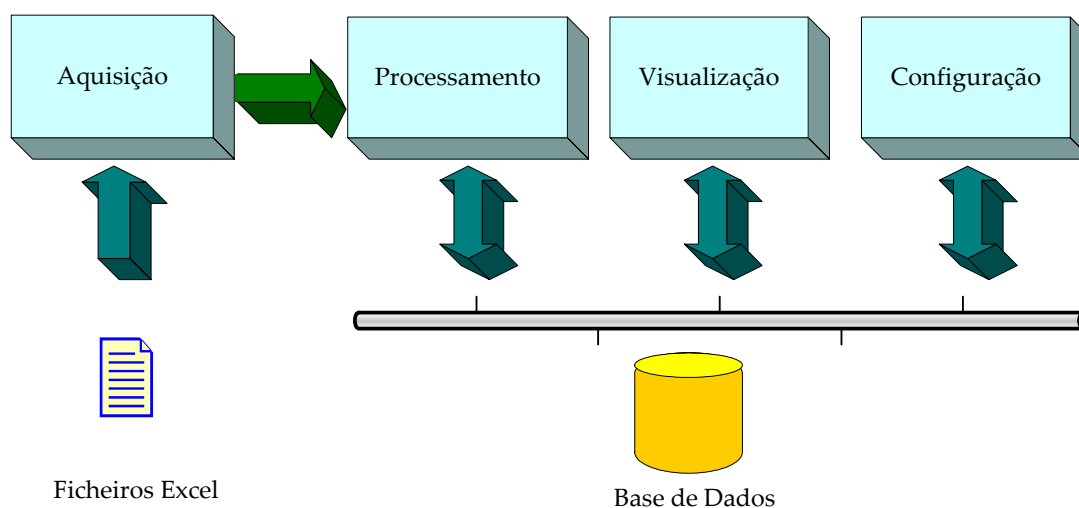


Figura 27 - Arquitectura implementada no desenvolvimento do sistema.

Através desta arquitectura, composta por vários módulos separados, torna-se possível o encaixe ou a substituição de diferentes módulos quando se tornar necessário. Esta particularidade simplifica possíveis alterações futuras ao sistema.

Para descrever a implementação do NeoScreen, vamos recorrer a pequenos diagramas de forma a simplificar a descrição das várias classes desenvolvidas para a sua implementação.

A aplicação NeoScreen foi desenvolvida com o Visual Studio.NET 2003, recorrendo ao C++ apoiado nas bibliotecas *Microsoft Foundation Class* (MFC). Estas bibliotecas impõem uma arquitectura pré-definida (Figura 28), que permite aos programadores com uma certa experiência em MFC, interpretar com facilidade qualquer aplicação desenvolvida em MFC. Fornecem também um conjunto classes que permitem uma diminuição muito significativa tempo empregue no desenvolvimento de aplicações.

A arquitectura MFC fornece automaticamente no momento da implementação quatro classes base, impondo desde a sua criação, um conjunto de regras bem definidas.

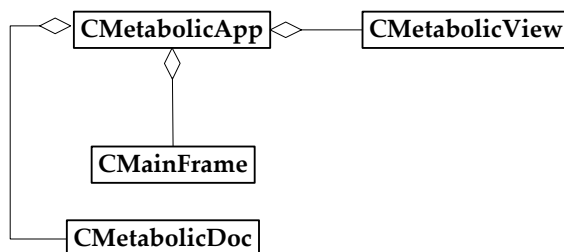


Figura 28 – Grupo de classes impostas pela arquitectura MFC.

As quatro classes impostas no momento da criação são:

- CMetabolicApp – classe a ser chamada no momento da execução da aplicação. Esta classe fica responsável por gerir todas as mensagens do sistema operativo e da aplicação;
- CMainFrame – Gere a parte visual respeitante ao menu, barra de atalhos, barra de status, etc;
- CMetabolicView – É responsável pela área visual central;
- CMetabolicDoc – Contém as estruturas de dados da aplicação e fica responsável por responder aos comandos de activação das caixas de diálogo gerais.

A classe CMetabolicApp, sendo a classe principal, tem agregadas as três classes impostas pela arquitetura MFC que é a CMetabolicDoc, a CMainFrame e a CMetabolicView.

Através deste modelo um programador com experiência em MFC, consegue ter uma visão geral de qualquer aplicação desenvolvida nesta arquitetura, sem necessidade de estar envolvido no desenvolvimento do código de uma determinada aplicação.

A classe CMainFrame, representada com outras classes pelo diagrama da Figura 29, tem agregada a classe CDlgServerConnect estando responsável pela ligação e autenticação do utilizador à base de dados. A classe CDadosIniciais, também agregada à classe CMainFrame, contém as propriedades de todos os parâmetros de configuração que um determinado utilizador tem quando está a trabalhar no NeoScreen. Esses parâmetros estão gravados na sua área pública no sistema operativo, não indo interferir com os parâmetros de outro utilizador ao trabalhar com o NeoScreen no mesmo sistema computacional. Esta concepção é implementada pela classe CSaveIniFiles, possibilitando a existência de parâmetros distintos para vários utilizadores no mesmo sistema computacional. A classe CIniFile é responsável pela gravação e leitura desses parâmetros em ficheiro.

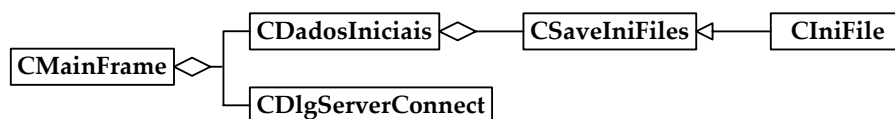


Figura 29 – Agregação de classes à classe CMainFrame.

O bloco Visualização é responsável pela visualização da informação estando definido através da classe CMetabolicView, contendo quatro classes agregadas que fornecem a visualização dos dados da aplicação NeoScreen, sendo elas a CLeftView, a CViewInformation, a CViewHistogram e a CViewPerson, representado pelo diagrama da Figura 30.

A classe CLeftView está responsável pela apresentação da árvore que suporta a maioria da informação acessível através do NeoScreen. Tem agregado a classe CDlgReport responsável pela geração de relatórios que se podem obter a partir de informação presente na árvore.

A classe CViewInformation é responsável por visualizar a informação adicional de um indivíduo, como as doenças, o diagnóstico, a informação da placa entre outras. A agregação das classes CInformation e CIdentification fornecem a estrutura de dados necessária ao suporte da informação. A CInformation é

responsável pelas doenças, diagnóstico, dados da placa, etc. A CIdentification tem a identificação do indivíduo em visualização.

A classe CDlgPrintInformation permite a impressão dos valores disponíveis na estrutura de dados da classe CInformation e CIdentification. A classe CDlgDiagnosisInfo permite a edição e alteração da identificação do indivíduo em análise. Tem agregada a classe CdlgDistricts para permitir a edição e alteração dos distritos.

A classe CViewHistogram permite a visualização dos histogramas que representam a evolução das concentrações do marcador seleccionado. A agregação da classe CDlgDataViewHistogram permite definir quais os intervalos temporais que se pretende visualizar.

A classe CViewPerson é responsável pela visualização das concentrações pertencentes ao indivíduo seleccionado no momento. Esta informação é visualizada em paralelo com a informação fornecida pela classe CViewInformation.

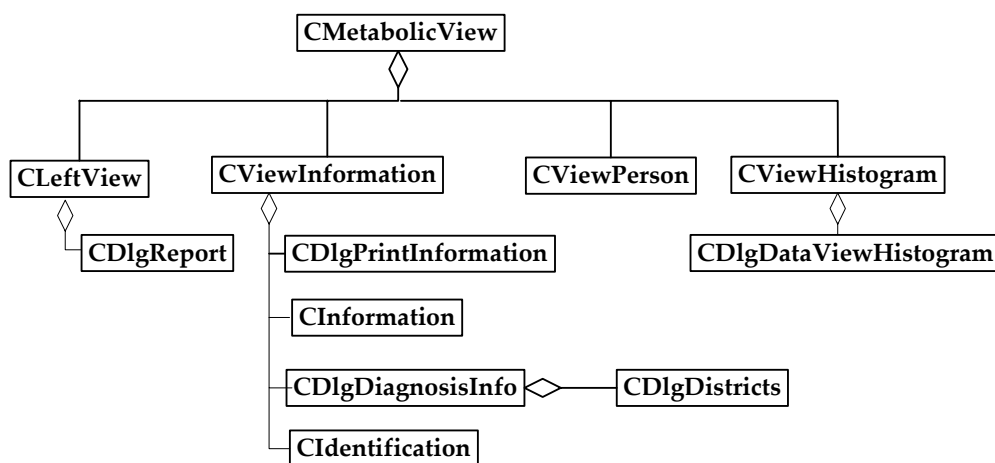


Figura 30 - Agregação de classes à classe CMetabolicView.

A classe CMetabolicDoc agrega dois tipos distintos de classes. As Classes que contêm a estrutura da dados e classes que permitem alterar, processar ou visualizar dados através de caixas de diálogo.

Na Figura 31 está representada a estrutura de dados que serve de suporte à aplicação NeoScreen. A estrutura está agregada à classe CMetabolicDoc.

A classe CMedia é responsável pela estrutura de dados que suporta os valores necessários para obter os histogramas da classe CViewHistogram. A classe

CMedia contém um vector dinâmico com objectos do tipo CvaloresMedia contendo todos os valores diários para o marcador em visualização.

A classe CDadosEstatisticos é responsável pelos dados estatísticos presentes na árvore. Esses dados são necessários à construção do relatório pela classe CDlgReport.

A classe CMarkers contém todos os marcadores presentes na base de dados sendo actualizada mediante qualquer alteração dos mesmos. Serve de referência para outras classes que necessitem de marcadores.

A classe CIndividuos é responsável por gerir os rastreios efectuados num determinado dia ou numa determinada consulta à base de dados. A classe CRastreios, agregada à classe CIndividuos, é um vector contendo todos os rastreios resultantes de uma consulta à base de dados. A classe CRastreios contém várias agregações como a CCalculateCompost necessária ao cálculo dos marcadores descritos como expressão matemática. A classe CMarkers contém os marcadores necessários para cada rastreio bem como as suas propriedades. A classe CArrDoenca é o vector responsável pelas doenças rastreadas ao indivíduo em questão. A CArrDoenca contém agregada a classe CArrMarkerDoenca sendo um vector de marcadores para a doença, assim como a classe CDoenca contendo todos os atributos da mesma.

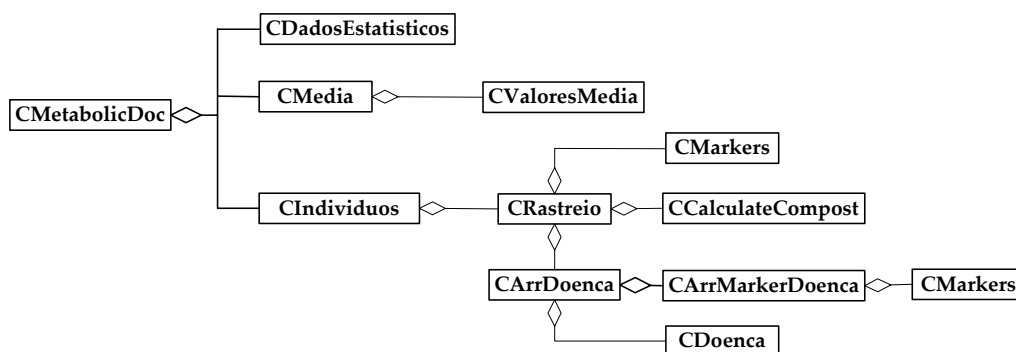


Figura 31 – Estrutura de dados implementada através de classes.

O bloco Configuração engloba todas as possibilidades de definição de parâmetros. As classes utilizadas para este fim são as classes que recorrem às caixas de diálogo, normalmente com o prefixo “CDlg”, estando representados no diagrama de blocos da Figura 32.

A classe CDlgSearch é responsável por gerir a caixa de diálogo na qual o utilizador define os parâmetros de consulta a efectuar na base de dados. Posteriormente, os indivíduos obtidos serão carregados numa estrutura de

dados disponibilizados pela classe CIndividuos, permitindo assim a visualização, alteração do diagnóstico e confirmação de doenças por parte do utilizador.

A classe CDlgUnits é responsável pela definição das unidades a atribuir aos marcadores.

A classe CDlgDefineMarker tem como objectivo a definição, alteração e eliminação de marcadores assim como as suas propriedades. A classe agregada CDlgComposeMarkers permite a definição das expressões matemáticas através dos marcadores definidos.

A classe CDlgRemoveFiles permite a eliminação de ficheiros existentes na base dados. A classe fornece uma lista em forma de árvore com todos os ficheiros submetidos à base de dados, organizados por ano, mes, dia, espectrómetro e nome da placa, permitindo assim a sua eliminação.

A classe CDlgDayReport proporciona uma consulta geral sobre um determinado dia escolhido pelo utilizador. São apresentados os indivíduos submetidos à base de dados no dia seleccionado, organizando-os por categorias. É possível também a criação de relatórios através da classe agregada CDlgReport.

A classe CDlgMachine permite a definição ou eliminação de espectrómetros. Permite também a definição de diagnósticos utilizando a mesma caixa de diálogo.

A classe CDlgStatisticsDay é responsável pela criação de estatísticas diárias representadas através de histogramas. É possível também gerar relatórios através da classe agregada CDlgReport. A classe CDlgStatisticsMonth e a classe CDlgStatisticsYear são responsáveis por gerar estatísticas referentes a meses e a anos, respectivamente.

A classe CDlgCDCCControl permite a definição, eliminação e alteração de controlos CDC. A classe agregada CControl contém a estrutura de dados necessária ao suporte da classe CDlgCDCCControl.

A classe CDlgTestControl executa os testes de monitorização das concentrações dos marcadores ao longo do tempo. Todos os dias monitoriza a evolução das concentrações para todos os marcadores. O alerta é lançado quando um determinado marcador tem um comportamento anormal.

A classe CdlgDisease é responsável pela criação, eliminação ou alteração de uma doença na base de dados. A classe agregada CDoenca fornece a estrutura de dados necessária ao suporte da classe CDlgDisease.

O bloco de Aquisição e o bloco Processamento são responsáveis pela leitura dos ficheiros e rastreio dos indivíduos provenientes da Aquisição, assim como a sua gravação em base de dados. Estes dois blocos estão definidos pela classe CProcessData assim como as suas agregadas. As classes CMarkers, CCalculateCompost, CNewIdentification, CSpreadSheet e CConjunto, agregadas ao CMetabolicDoc, fornecem todas as ferramentas necessárias ao processamento da informação importada através dos ficheiros. A classe CNewIdentification permite a identificação da placa quando não é possível obtê-la através do ficheiro de entrada. A classe CSpreadSheet lê toda a informação contida no ficheiro para uma estrutura de dados presente na classe. A classe CConjunto contém a regra que será confrontada com os dados lidos de forma a realizar os rastreios. Também é possível produzir relatórios através da classe CDlgReport.

A classe CDlgUpdate é responsável pelo cálculo diário de todos os percentis definidos nos marcadores. A classe agregada CTestaCalculoGeral realiza o controlo diário para calcular os percentis.

A classe CDlgRegra permite eliminar, definir e alterar regras. Necessita da classe agregada CConjunto para fornecer uma estrutura de dados de suporte para a regra em edição. Se alguma alteração for efectuada que afecte os cálculos dos percentis, é necessário proceder a novos cálculos através da classe CDlgUpdate.

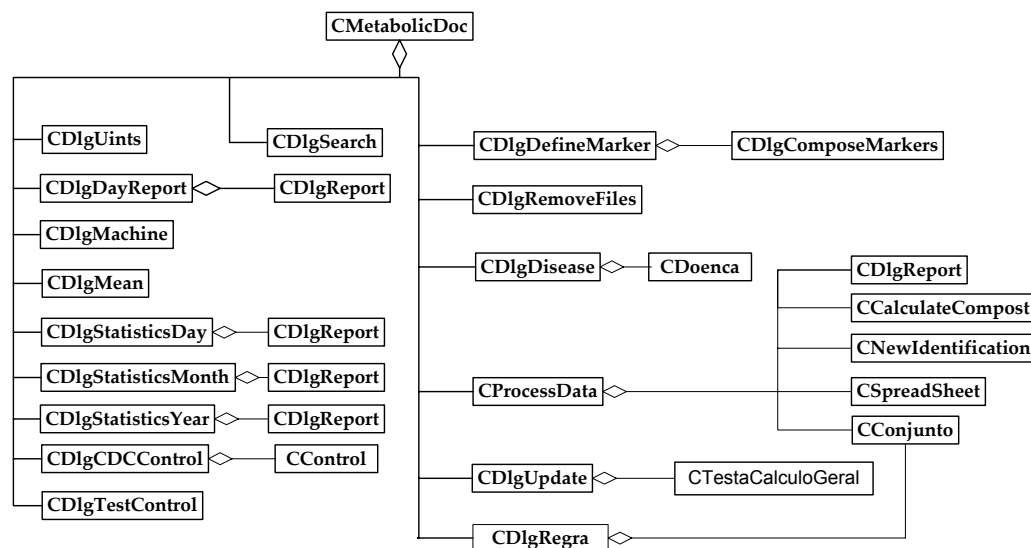


Figura 32 – Classes agregadas à classe CMetabolicDoc.

5.4 NeoScreen – Interface de utilizador

A interface do utilizador foi concebida de forma a simplificar a análise e a dar total confiança aos especialistas em diagnóstico, tipicamente cépticos e cautelosos relativamente a uma aplicação que realiza diagnósticos. A aplicação NeoScreen foi desenvolvida a pensar no paradigma do explorador de ficheiros de forma a tornar mais fácil a interacção com o utilizador. A Figura 33 é a janela principal da aplicação. No lado direito consta uma árvore onde estão organizados os grupos das várias entidades que podem ser consultadas. O resultado da consulta é apresentado no lado direito da janela. Através do botão assinalado com uma elipse verde poderão efectuar-se variadas consultas na base de dados, como a procura por indivíduo, data, espectrómetro, doença, concentração ou diagnóstico. Os indivíduos resultantes da procura são organizados por categorias definidas pelas especificações da aplicação. Através da selecção de um determinado indivíduo presente numa determinada categoria, obtêm-se variadas informações como as suas concentrações ou doenças suspeitas. O utilizador poderá então introduzir anotações como a confirmação de doenças, diagnósticos e a identificação do indivíduo em causa. Poderá ainda imprimir o resultado ou adicionar observações que ache prementes para o caso em estudo.

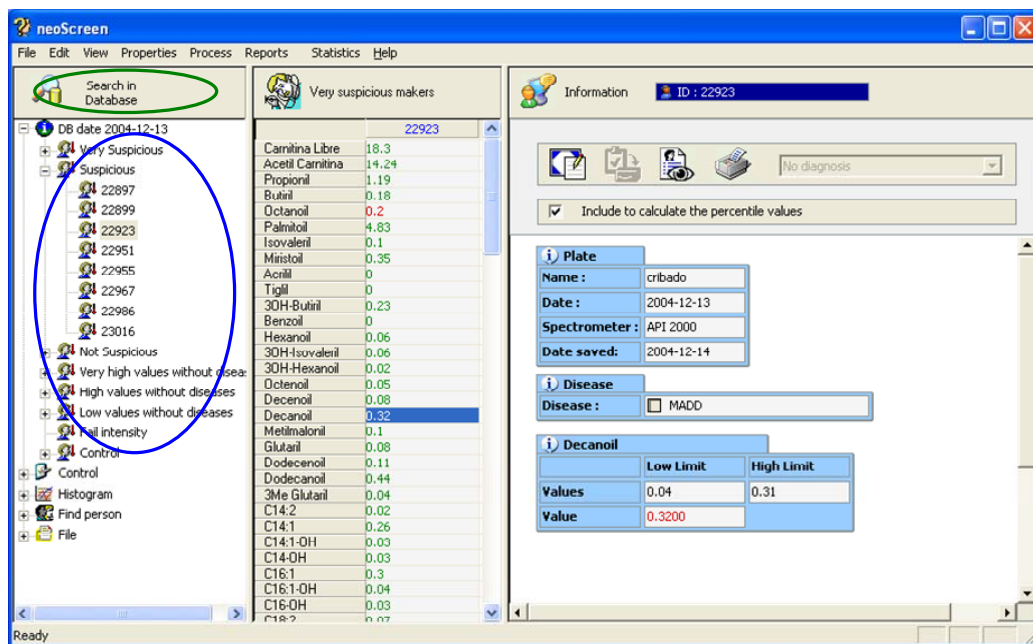


Figura 33 – Janela principal da aplicação NeoScreen.

O item “control” e o item “histogram” presentes na árvore dão acesso às cartas de controlo da Figura 34. As cartas de controlo têm a função de permitir visualizar a evolução das concentrações dos marcadores ao longo do tempo.

Cada ponto corresponde a uma média das concentrações inseridos na base de dados, para um determinado dia e marcador. No entanto, só os marcadores que pertençam aos indivíduos inseridos em categorias definidas para o cálculo das médias é que entram para o cálculo da média. As categorias que servem de referência para os cálculos são definidas pelo utilizador.

Posteriormente é calculada a média global de todas as médias diárias assim como o desvio padrão. A linha central corresponde à média global, as duas linhas mais próximas da linha central correspondem a duas vezes o desvio padrão e as duas linhas mais distantes da linha central correspondem a três vezes o desvio padrão. O utilizador poderá ainda incluir ou excluir uma determinada média diária do cálculo da média global, bastando para tal, “picar” em cima do ponto. Com esta propriedade é possível eliminar valores que introduzam ruído à média global.

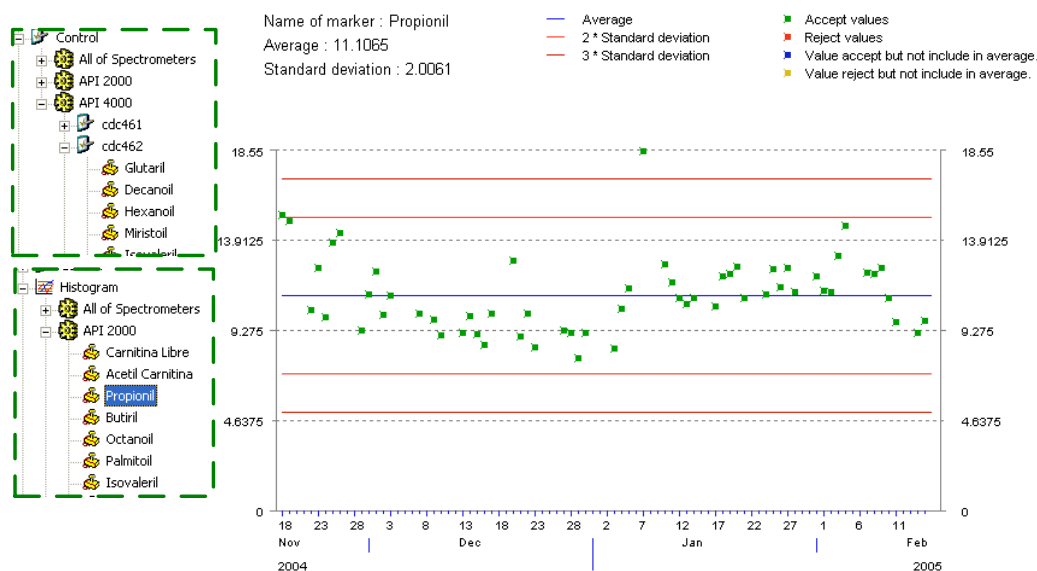


Figura 34 – Cartas de controlo para monitorizar a evolução das concentrações.

As cartas de controlo são construídas para os controlos CDC e para os marcadores em geral. No caso dos controlos CDC as cartas são organizadas por espectrómetro, controlo CDC e posteriormente os marcadores que correspondem ao controlo. No caso dos marcadores em geral, as cartas de controlo estão inseridas no item “histogram” e organizadas por espectrómetro e marcadores.

A importação dos indivíduos para a aplicação é efectuada através da leitura dos ficheiros Excel devidamente formatados. O utilizador poderá no entanto gravar a informação em base de dados, ou efectuar somente o rastreio, para visualizar o resultado da análise, sem necessidade de proceder à gravação.

A Figura 35 corresponde a uma primeira análise efectuada a um ficheiro lido através da aplicação. Esta primeira análise contém os marcadores reconhecidos e rejeitados, os indivíduos aceites e com falhas nos marcadores de intensidade e a identificação dos controlos CDC. O rastreio é efectuada à posteriori, ocorrendo no momento da gravação em base de dados ou efectuada no momento de carregar os indivíduos somente para análise. O rastreio é realizado confrontando os marcadores dos indivíduos lidos com o padrão escolhido pelo utilizador.

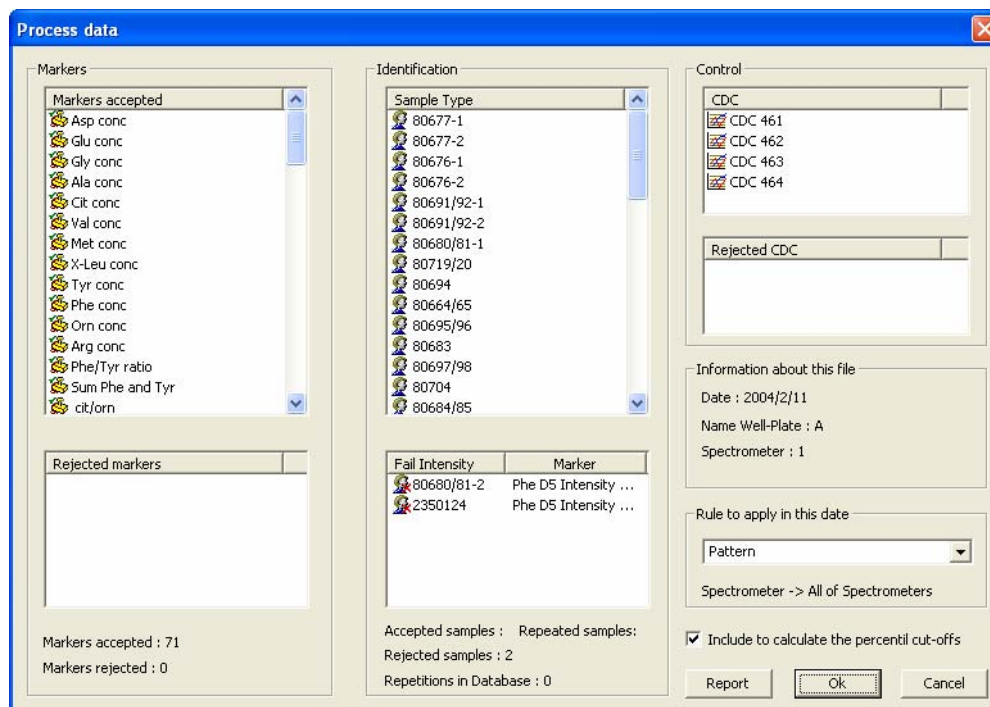


Figura 35 – Processo de leitura e rastreio dos indivíduos através da aplicação NeoScreen.

O utilizador terá numa primeira fase de definir todos os marcadores presentes nos ficheiros que irá importar para a base de dados. A Figura 36 mostra os marcadores presentes na base de dados, podendo no entanto efectuar alterações aos marcadores presentes.

Cada parâmetro tem quatro propriedades:

- Nível de controlo - indica que será o utilizador a definir qual o limite superior e inferior do marcador.
- Percentil - indica que será a aplicação a calcular os limites superior e inferior para determinado marcador. A aplicação todos os dias calcula os percentis para todos os marcadores definidos como tais, baseando-se em concentrações passadas e validadas pelo utilizador;

- Intensidade - indica que determinado marcador tem como função controlar concentrações pré-definidas introduzidas nas amostras a analisar. O utilizador define um limite mínimo para determinado marcador. Se esse limite não for alcançado pelo marcador o utilizador terá que repetir a análise feita ao indivíduo no MS/MS;
- Unidade - indica qual a unidade para determinado marcador.

O utilizador poderá ainda definir expressões matemáticas a partir dos parâmetros existentes recorrendo à tecla "Compose". A aplicação calcula automaticamente as concentrações para esses parâmetros.

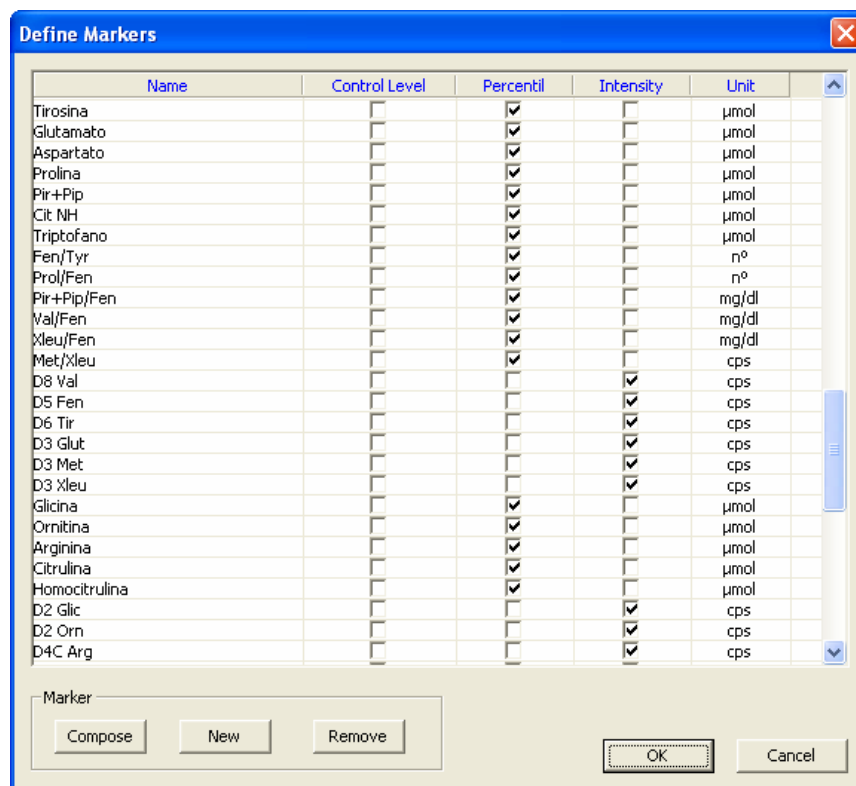


Figura 36 - Definição dos marcadores.

A aplicação está construída de modo a suportar mais do que um espectrómetro de massa. O utilizador terá que definir pelo menos um espectrómetro de massa como mostra a Figura 37. Por omissão, a aplicação tem o "All of Spectrometers" definido, mas o utilizador terá que definir obrigatoriamente um espectrómetro para além deste.

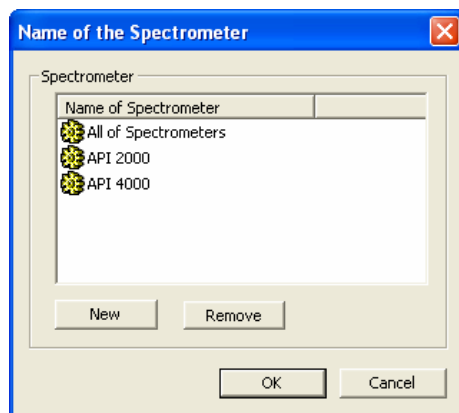


Figura 37 - Definição de espectrómetros de massa.

A Figura 38 contém a caixa de diálogo onde se definem os controlos CDC. A cada controlo corresponde um determinado conjunto de marcadores definidos pelo CDC. Os nomes dos controlos CDC terão de ser definidos previamente para que não sejam considerados indivíduos no momento da leitura dos ficheiros.

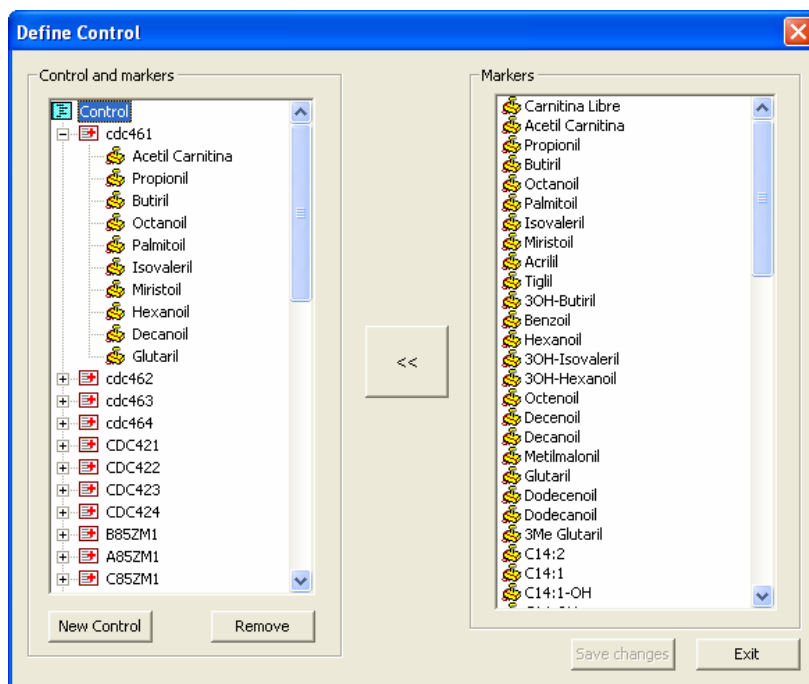


Figura 38 - Definição de controlos CDC.

Na janela de diálogo Figura 39 o utilizador poderá definir as doenças passíveis de rastreio através da aplicação NeoScreen. O utilizador ao criar uma doença terá de indicar quais os marcadores associados assim como os limites para cada marcador. Neste processo, o utilizador pode indicar se determinado marcador é

activado somente quando existe uma concentração que passa acima do limite, abaixo do limite ou ambos. Uma determinada doença só é apresentada como suspeita num indivíduo se todos os marcadores indicados na doença estiverem fora dos limites definidos.

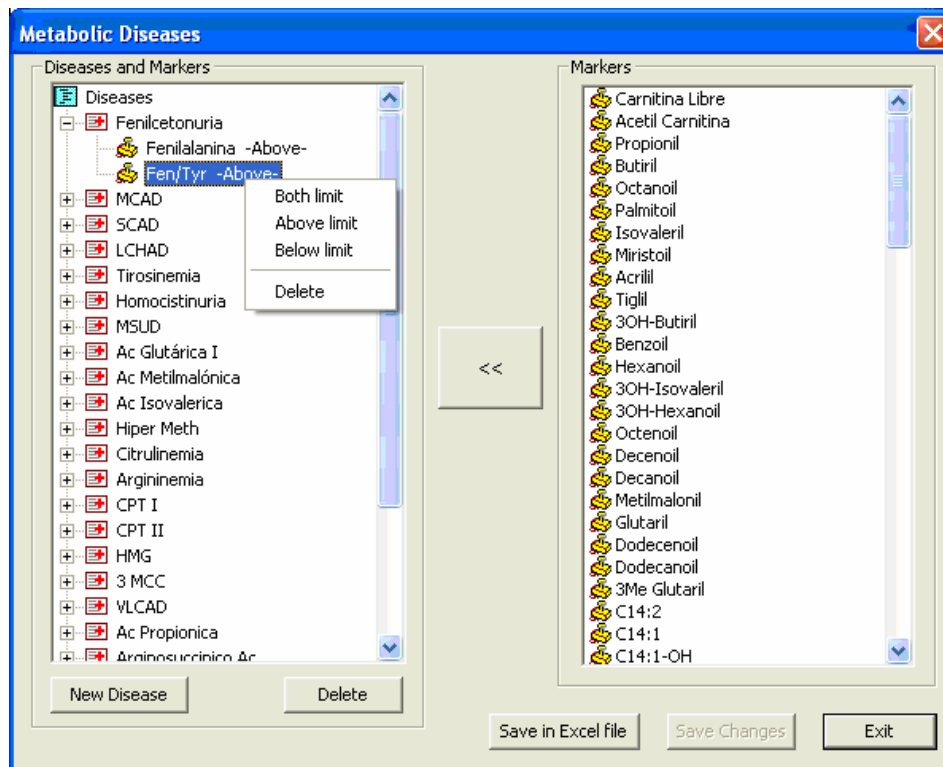


Figura 39 - Definição doenças.

A Figura 40 contém a caixa de diálogo onde o utilizador poderá criar ou alterar padrões. Os padrões serão usados para efectuar os rastreios confrontando os valores definidos nas mesmas com os valores lidos dos ficheiros Excel. Cada padrão definido pelo utilizador, contém todos os marcadores definidos no sistema, assim como os seus limites que irão ser confrontados com os marcadores dos recém-nascidos.

O utilizador tem a possibilidade de definir os limites dos marcadores, caso tenham como propriedade 'Control level', ou definir os percentis caso tenham a propriedade 'Percentil'.

Na caixa com o nome 'Parameters of the percentil values' o utilizador pode definir qual o intervalo de tempo que serve de referência para o cálculo dos percentis. Pode também excluir os valores que estejam fora do valor definido pelo percentil, eliminando ruído introduzido pelas amostras que têm concentrações fora do normal.

Os valores que servem de referência para o cálculo do percentil podem ser escolhidos pelo utilizador, ou escolhidas pela aplicação através das classes a que os indivíduos pertencem. Essa definição pode ser definida na caixa 'Values of reference to calculate the percentil'.

Figura 40 - Definição de regras.

A aplicação contém uma ferramenta que permite uma vigilância permanente sobre a evolução das concentrações admitidas na base de dados. Quando um ficheiro é submetido à base de dados, a aplicação testa se as médias diárias correspondem a algum padrão de anormalidade definido nas especificações da aplicação. Na Figura 41 podemos ver que o marcador "Fen MRM" falhou o teste de três desvios padrões. Quer isto dizer que, o marcador tem um valor superior ou inferior à média global mais três desvios padrões. Este resultado poderá ser confirmado através das cartas de controlo fornecidas pela aplicação.

Este teste é feito automaticamente aquando da leitura de um ficheiro. O utilizador tem ainda a possibilidade de executá-lo manualmente, para um melhor controlo dos marcadores.

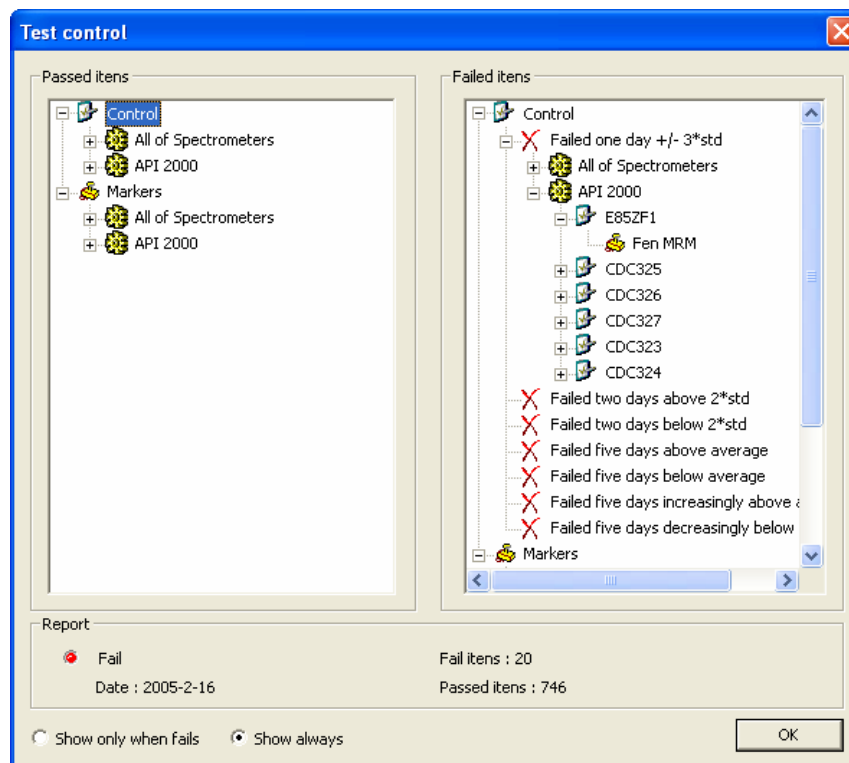


Figura 41 - Controlo dos marcadores através das médias diárias.

Na Figura 42 consta uma janela de diálogo na qual o utilizador poderá efectuar variadas estatísticas obtendo o resultado através de um histograma. É possível recorrer a outras caixas de diálogo para efectuar estatísticas por dia ou ano.

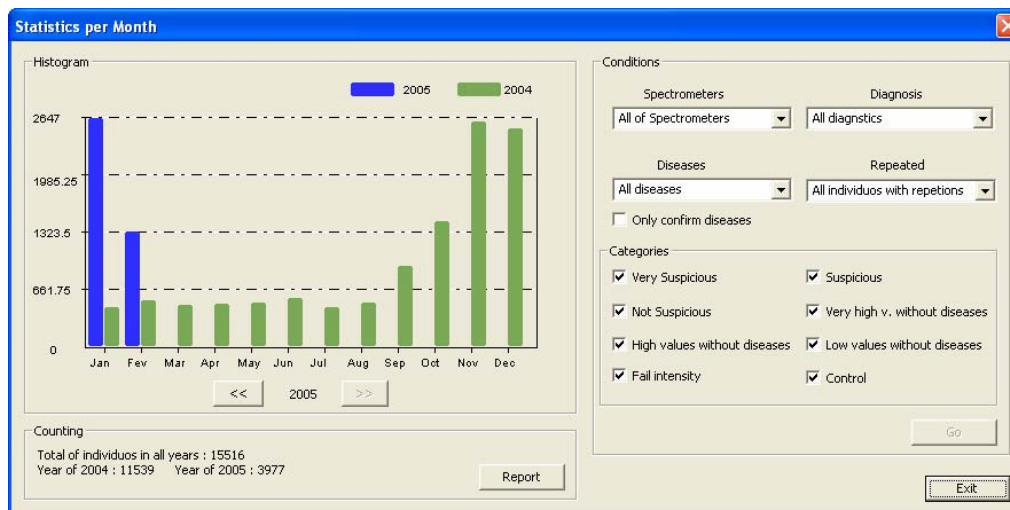


Figura 42 – Estatísticas obtidas através dos histogramas.

A aplicação está construída para ser um sistema multiposto como mostra a Figura 43. Com esta propriedade, os utilizadores, devidamente registados na base de dados, poderão efectuar análises e rastreios em simultâneo.

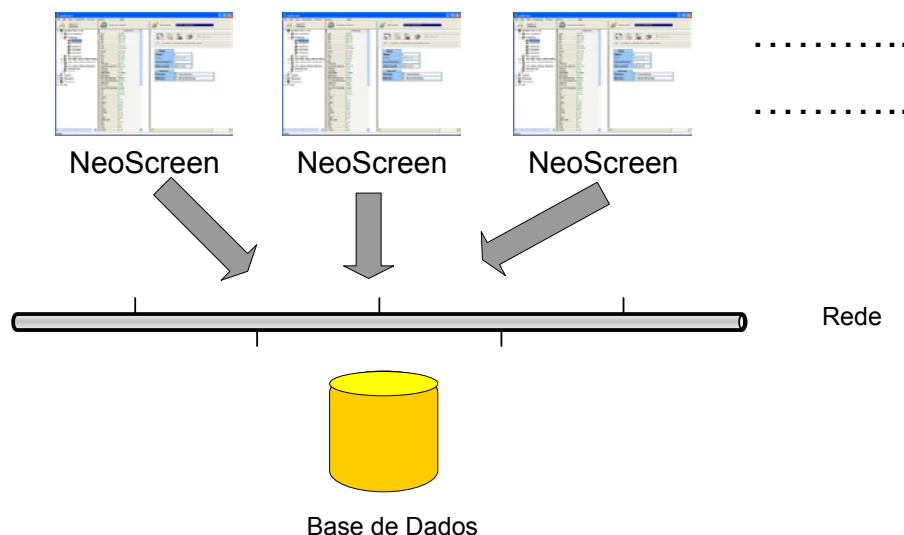


Figura 43 – Multiposto através do NeoScreen.

5.5 Sumário

O NeoScreen foi a aplicação desenvolvida para preencher os requisitos traçados pelo IGM. A aplicação importa dados através de ficheiros, efectuando posteriormente o rastreio e armazenamento de toda a informação numa base de dados. Em qualquer altura o utilizador poderá consultar a informação analisada bem como ter acesso ao histórico através de um conjunto de ferramentas fornecidas pelo NeoScreen.

No desenvolvimento da aplicação privilegiou-se a interface do utilizador, resultado de inúmeras interações com os especialistas do IGM no decorrer do desenvolvimento. A aplicação proporciona o tratamento dos dados obtidos através do MS/MS, de forma rápida e intuitiva, diminuindo, em muito, o tempo de rastreio. Devido a estes atributos, a possibilidade de existirem falhas no rastreio é minimizada.

Futuramente, através do repositório de informação obtido ao longo do tempo, é possível efectuar estudos longitudinais, permitindo traçar padrões entre regiões, ou grupos de pessoas da mesma família.

6 Resultados

A primeira instalação do NeoScreen ocorreu em Setembro de 2004, tendo-se verificado que existia a necessidade de realizar algumas alterações ao software. Essas alterações permitiram melhorar as análises dos resultados e também algumas especificações que foram surgindo no decorrer do trabalho. Uma nova versão surgiu em Novembro de 2004 estando em testes até Março de 2005. A partir de Março, arrancou o projecto experimental abrangendo cerca de metade do território nacional, mais concretamente até ao distrito de Coimbra.

Em Outubro de 2004 surgiu a oportunidade de instalar uma versão do NeoScreen no Hospital Clínico Universitário de Santiago (HCUS), sendo a entidade responsável pelo rastreio neo-natal na província espanhola da Galiza. O número de rastreios é muito inferior aos rastreios efectuados em Portugal, rondando 100 recém-nascidos por dia. Este contacto foi estabelecido devido às boas relações entre IGM e HCUS. Até ao momento o HCUS realizava todo o rastreio neo-natal de forma manual, devido ao número de nados vivos ser pequeno. No entanto, o HCUS decidiu utilizar a aplicação para introduzir um controlo mais rigoroso no rastreio neo-natal.

Actualmente já foram submetidos à base de dados um total de 15.421 indivíduos no HCUS e de 11.916 indivíduos no IGM. A razão para o HCUS ter submetido mais valores do que o IGM, mesmo com menos recém-nascidos/dia, prende-se com o facto de o HCUS ter inicializado os testes com o recurso ao MS/MS em Janeiro de 2004. Contendo, toda essa informação em ficheiros Excel, foi submetendo todos os ficheiros até à presente data. Convém também referir, que o rastreio de recém-nascidos, através da tecnologia MS/MS, ainda não está estendido a todo o território nacional.

Os dados introduzidos, assim como a sua constante utilização, permitiram certificar a robustez da aplicação desenvolvida, servindo também de base a algumas estatísticas apresentadas neste capítulo.

Os dados obtidos são referentes ao período compreendido entre Novembro de 2004 e Março de 2005 e encontram-se na Tabela 8. A tabela compreende duas categorias de valores, sendo elas:

- Percentil – valores calculados através da aplicação NeoScreen;
- Empírico – valores definidos pelo laboratório, obtidos através de literatura especializada.

Ao confrontar os calculados através dos percentis com os empíricos podemos constatar que existem pequenas variações entre os limites. Mas, são estas pequenas variações que vão permitir diminuir o número de falsos positivos.

Tabela 8 – Valores calculados através dos percentis e valores empíricos do IGM. Valores em $\mu\text{mol/L}$.

IGM						
Marcadores	Percentil				Empírico	
	Baixo		Alto		Baixo	Alto
	2,5	5	95	99,5		
Phe	27,79	31,48	86,51	125,7	20	130
Ala	155,89	182,58	710,59	1207,37	164	819
Arg	1,07	1,45	13,07	23,78	1	32
Orn	29,53	35,86	261,9	459,33	25	500
Xleu	90,16	102,14	253,7	332,89	72	350
Tyr	43,66	54,37	284,11	504,55	58	400
Phe/Tyr	0,15	0,19	0,86	1,86	0	2
Val	76,91	87,25	250,76	358,98	77	417
Cit	4,37	4,91	17,09	36,78	4	29
C0	10,46	12,46	44,22	73,74	11	56
C2	8,15	9,38	33,66	50,87	12	43
C3	0,5	0,64	3,23	5,38	0,74	5,46
C5	0,06	0,07	0,44	2,12	0,06	1,19
C6	0	0	0,08	0,16	0	0,13
C5-OH	0,06	0,07	0,25	0,4	0,06	0,5
C8	0	0,01	0,12	0,24	0	0,26
C10	0,02	0,04	0,18	0,35	0	0,39
C10:1	0,05	0,07	0,79	1,67	0	0,14
C14	0,08	0,1	0,34	0,55	0,4	0,52
C14:1	0,03	0,04	0,16	0,29	0,02	0,28
C16	0,09	1,15	4,83	7,17	0,05	7,4
C16-OH	0	0,01	0,05	0,08	0	0,08
C18	0,37	0,45	1,59	2,31	0,25	2,3
C18-OH	0,01	0,01	0,05	0,08	0	0,05
C18:1	0,6	0,76	2,45	3,43	0,5	3,6
C18:1-OH	0,01	0,01	0,05	0,08	0	0,06

Na Tabela 9 podemos consultar os valores obtidos através dos dados do HCUS. Podemos desde já ver algo interessante. O HCUS admite todos os recém-nascidos para calcular os valores dos percentis, originando valores muito distantes em relação aos valores empíricos, ou em relação aos valores obtidos no IGM. Contudo, os percentis escolhidos para o cálculo dos mesmos são mais dilatados no caso do HCUS. Porém, o IGM definiu que só os recém-nascidos considerados normais, que contêm todas as concentrações dentro dos limites, são admitidos para cálculo dos percentis. Com esta opção, não é adicionado erro nos cálculos dos percentis, introduzido através de recém-nascidos suspeitos de doença, obtendo assim valores mais próximos dos valores empíricos.

Tabela 9 - Valores calculados através dos percentis e valores empíricos do HCUS. Valores em $\mu\text{mol/L}$.

Marcadores	HCUS					
	Percentil				Empírico	
	Baixo		Alto		Baixo	Alto
	0,5	1	99	99,5		
Phe	23,08	25,25	453,73	562,57	23,5	141
Ala	132,76	149,02	1010,49	1116,42	153,95	533,6
Arg	2,21	2,37	26,86	31,98	1,95	17,13
Orn	15,07	17,11	261,37	312,554	17,88	150,2
Xleu	97,88	112,72	833,28	885,92	125,03	416,61
Tyr	14,57	20,16	260,76	291,43	25,58	245,08
Phe/Tyr	0,16	0,19	8,15	11,82	0,19	1,91
Val	63,74	73,77	305,35	327,99	79,22	256,56
Cit	5,086	5,43	38,36	47,96	4,88	28,27
C0	7,29	8,21	56,39	61,12	7,7	42,04
C2	4,66	5,487	48,54	52,11	7,54	51,21
C3	0,34	0,39	4,82	5,342	0,46	4,82
C5	0,03	0,03	0,81	1	0,04	0,81
C6	0	0,01	0,22	0,25	0,01	0,22
C5-OH	0,04	0,04	0,330	0,380	0,06	0,31
C8	0	0,01	0,19	0,21	0,01	0,19
C10	0,01	0,02	0,34	0,39	0,04	0,31
C10:1	0,01	0,01	0,21	0,224	0,01	0,21
C14	0	0	0,026	0,03	0	0,02
C14:1	0,02	0,03	0,41	0,47	0,02	0,29
C16	0	0	0,42	0,48	0	0,05
C16-OH	0	0	0,06	0,07	0	0,06
C18	0,14	0,18	1,88	1,99	0,24	1,9
C18-OH	0	0	0,04	0,05	0	0,04
C18:1	0,26	0,34	3,02	3,18	0,42	2,96
C18:1-OH	0	0	0,7	0,076	0	0,6

É de evidenciar que o HCUS utiliza a aplicação, mais para monitorizar a evolução dos parâmetros ao longo do tempo, do que, para efectuar diagnósticos. Isto deve-se ao facto de o número de recém-nascidos ser relativamente baixo.

A monitorização dos marcadores é realizada através da análise das cartas de controlo como mostra a Figura 44. A aplicação permite a análise automática dos padrões definidos como anormais. O utilizador poderá, posteriormente, confirmar as anomalias através da consulta das cartas de controlo. Tem também a possibilidade de eliminar algum valor que não tenha interesse, permitindo assim, a eliminação do ruído introduzido no cálculo da média.

Os padrões anormais estão assinalados por elipses de várias cores (números) na Figura 44. A elipse azul (1) contém cinco ou mais dias seguidos abaixo da média. A elipse verde (2) tem dois valores entre duas vezes o desvio padrão e três vezes o desvio padrão. A elipse vermelha (3) contém cinco valores acima da

média que evoluem progressivamente. A elipse laranja (4) contém um ou mais valores acima de três vezes do desvio padrão.

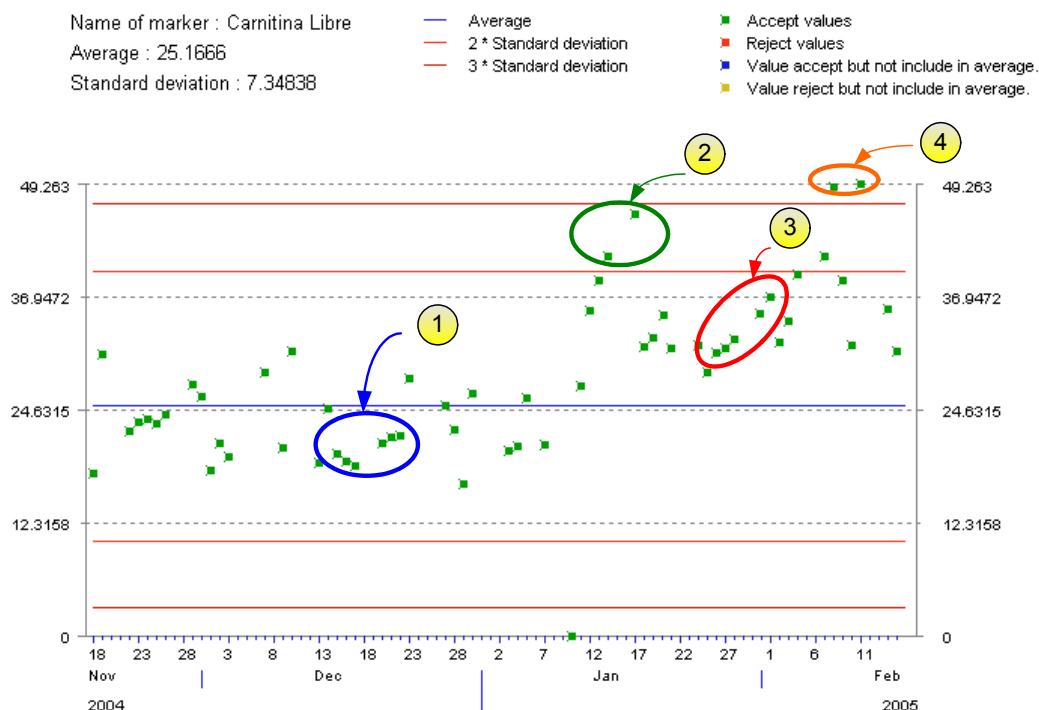


Figura 44 – Análise de uma carta de controlo.

Estes padrões, considerados anormais, poderão indicar que algum procedimento na preparação da amostra ou algum problema de calibração no equipamento poderá estar a ocorrer. A aplicação, ao detectar alguma destas anomalias assinaladas na figura, gera automaticamente o alerta, libertando o investigador de uma diária e constante visualização das cartas de controlo à procura de anomalias.

O IGM já detectou quatro casos confirmados de novas doenças através da aplicação NeoScreen, foram eles:

- 2 MCAD – esta patologia enquadra-se no grupo das acidúrias orgânicas e fica a dever-se à elevada concentração das acilcarnitinas C6, C8, C10, C10:1 presentes no sangue;
- 1 MSUD – esta patologia enquadra-se no grupo das aminoacidopatias e fica a dever-se à elevada concentração dos aminoácidos Leu/Ile e Val;
- 1 3HMG – esta patologia enquadra-se no grupo das acidúrias orgânicas e fica a dever-se à elevada concentração da acilcarnitina C5-OH.

Como se pode ver na Figura 45, no espaço compreendido entre Janeiro e Março de 2005 já se evidencia a importância de recorrer à nova tecnologia, detectando

mais quatro casos de doenças que não tinha sido possível sem a ajuda do NeoScreen. Graças às novas técnicas de rastreio juntamente com sistemas de apoio à decisão, como é o caso do NeoScreen foi possível melhorar o nível de vida de quatro crianças num curto espaço de tempo, que corresponde de Janeiro a Março de 2005.

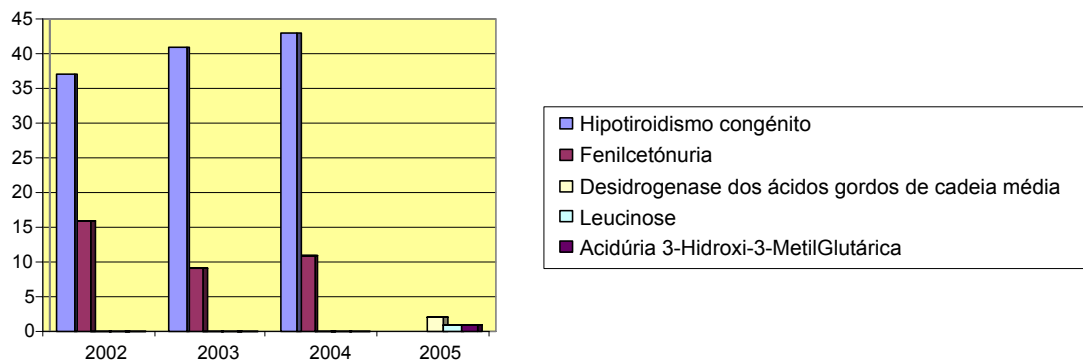


Figura 45 – Evolução dos recém-nascidos identificados com patologias.

Até ao momento a aplicação NeoScreen detectou todos os recém-nascidos com doenças metabólicas, sendo aparentemente zero o número de falsos negativos. Escreve-se “aparentemente” devido à não existência de casos em que a doença se manifestou sem que o IGM tenha detectado a falha.

Com estes resultados, o NeoScreen cumpre inteiramente a tarefa para qual foi desenhado, possibilitando ainda uma fácil parametrização, tornando-o totalmente dinâmico, servindo ainda de suporte para possíveis estudos futuros que se puderam realizar graças à base de dados que suporta toda a informação.

7 Conclusões e trabalho futuro

Ao longo desta tese foram apresentados os métodos envolvidos no rastreio neonatal, bem como as dificuldades que os investigadores têm no tratamento de grandes quantidades de informação gerados pelos novos métodos, que permitem aumentar o número de doenças rastreáveis.

O objectivo proposto seria desenvolver um sistema de informação que permitia tratar e organizar toda a informação gerada pelo sistema de análise, possibilitando o acesso à informação, efectuar rastreios e consultas de forma rápida seu rastreio

Com o NeoScreen os laboratórios de rastreio conseguem uma gestão permanente da informação gerada pelo MS/MS, diminuindo o tempo e erros de análise, conseguindo aumentar em muito o número de rastreios efectuados diariamente, aumentando também a sua qualidade.

A nível visual, o sistema assenta em paradigmas familiares, permitindo uma fácil utilização e consulta por parte do investigador. Os dados são importados para o NeoScreen através da leitura de ficheiros contendo as concentrações dos respectivos marcadores. Após o diagnóstico realizado pelo sistema, a informação será gravada em base de dados para posteriores consultas e análises.

Durante o processo de rastreio a aplicação separa os indivíduos por diferentes categorias. Esta separação permite ao utilizador dar especial atenção às categorias de suspeito e muito suspeito.

O sistema NeoScreen fornece um conjunto de ferramentas matemáticas e estatísticas que permitem calcular os limites dos marcadores necessários à activação das doenças potencialmente suspeitas.

É um sistema totalmente dinâmico permitindo configurações de vários parâmetros como a definição de novos marcadores, novas doenças, percentis e seus limites, desenvolvimento de expressões matemáticas sobre marcadores, entre outros.

O sistema também mantém um controlo efectivo das concentrações dos marcadores, efectuando uma monitorização diária à média e ao desvio padrão. Se existir algum desvio significativo, o sistema emite uma alerta permitindo a verificação da falha.

Futuramente, e com o crescimento da base de dados, poderão adicionar-se novas ferramentas estatísticas, ou integrarem-se ferramentas de exploração de dados (mineração, por exemplo), que permita recolher padrões não observados até ao momento, dada à fase inicial do programa.

A base de dados também permitirá realizar estudos epidemiológicos a nível nacional, ou efectuar estudos longitudinais, possibilitando traçar padrões de predominância de doenças em diferentes regiões.

A aplicação tem revelado bom desempenho, boa robustez, permitindo assim um rastreio rápido e com um controlo efectivo dos valores a analisar. No entanto, tem-se verificado que algumas alterações serão necessárias para melhorar a produtividade da aplicação, assim como, a criação de relatórios diários, possibilidade de gerar listagens sobre os rastreios efectuados, permitir visualizar todas os marcadores.

O projecto experimental arrancará em Março de 2005 abrangendo cerca de metade do território nacional, primeiro a região do Grande Porto e posteriormente até ao distrito de Coimbra. Se tudo correr como previsto, no prazo de um ano o rastreio de novas doenças poderá estender-se a todo território nacional.

8 Anexos

Anexo 1 – Acrónimos dos aminoácidos.

Tabela 10 – Aminoácidos e suas abreviaturas

Aminoácido	Abreviatura (3 letras)	Abreviatura (1 letra)
Acido aspártico	Asp	D
Acido Glutámico	Glu	E
Arginina	Arg	R
Lisina	Lys	K
Asparagina	Asn	N
Histidina	His	H
Glutamina	Gln	Q
Serina	Ser	S
Treonina	Thr	T
Alanina	Ala	A
Glicina	Gly	G
Valina	Val	V
Prolina	Pro	P
Leucina	Leu	L
Fenilalanina	Phe	F
Tirosina	Tyr	Y
Isoleucina	Ile	I
Metionina	Met	M
Triptofano	Trp	W
Cisteína	Cys	C
Final	Stop	

Anexo 2 – Descrição das doenças rastreadas em Portugal através da tecnologia MS/MS

São actualmente 13 doenças actualmente rastreáveis em Portugal com o auxílio do NeoScreen:

- Fenilcetonúria;
- Leucinose;
- Citrulinemia;
- Acidúria ArgininoSuccínica
- Acidúria Propiónica
- Acidúria Metilmalónica
- Acidúria Isovalérica
- Acidúria 3-Hidroxi-3-MetilGlutárica
- Acidúria Glutárica Tipo I
- Desidrogenase dos Ácidos Gordos de Cadeia Média
- Desidrogenase de 3- Hidroxi-AcilCoA de Cadeia Longa
- Desidrogenase dos Ácidos Gordos de Cadeia Muito Longa
- Carnitina-Palmitoil Transferases I

A **Fenilcetonúria** (PKU) é uma doença hereditária do metabolismo, causada pelo excesso do aminoácido fenilalanina, atacando principalmente o cérebro e causando deficiência mental.

Devido a uma mutação do gene PAH no cromossoma 12 ocorre uma deficiência na enzima fenilalanina hidroxilase (PAH). Esta enzima é responsável pela metabolização da fenilalanina em tirosina, tendo como consequência directa a acumulação do aminoácido em quantidades tóxicas para o organismo [19].

O tratamento desta doença consiste numa dieta hipoproteica restrita, especialmente à base de leite e de alimentos que não contenham fenilalanina.

Actualmente também se considera o rácio da fenilalanina com a tirosina para proceder ao rastreio da doença. Recorrendo a este marcador consegue-se diminuir o numero de falsos positivos no momento do rastreio [63].

A **Leucinose** (MSUD) é uma doença hereditária do metabolismo dos aminoácidos ramificados: Leucina, Isoleucina e Valina.

Devido à deficiência no complexo multienzimático da desidrogenase dos ácidos α -cetónicos de cadeia ramificada (BCKAD), estes aminoácidos e alguns compostos deles derivados acumulam-se em quantidades tóxicas para o organismo.

O tratamento desta doença consiste numa dieta hipoproteica restrita nos aminoácidos leucina, isoleucina e valina e suplementada com carnitina.

A **Citrulinemia** (CIT) é uma doença hereditária do metabolismo devida à deficiência na enzima Argininosuccinato Sintetase (AS). Esta enzima actua no ciclo da ureia, um importante mecanismo envolvido celular de eliminação da amónia. O ciclo da ureia envolve uma série de passos bioquímicos em que o nitrogénio, um resíduo produzido pelo metabolismo das proteínas, é removido do sangue e convertido em ureia. Normalmente, a ureia é transferida para a urina e removida do organismo. Nas desordens que envolvem o ciclo da ureia, o nitrogénio é acumulado no organismo em forma de amónia, sendo uma substância tóxica e existindo dificuldade na sua remoção do organismo. Este composto forma-se durante o catabolismo celular normal dos aminoácidos, mas em elevadas concentrações é tóxico para as células.

A Citrulinemia caracteriza-se pela acumulação e excreção na urina de quantidades muito elevadas de citrulina e amónia.

O tratamento desta doença consiste numa dieta hipoproteica suplementada com arginina e carnitina e na administração de fenilbutirato/ benzoato de sódio para facilitar a eliminação da amónia.

A **acidúria ArgininoSuccínica** (ASA) é uma doença hereditária do metabolismo devida à deficiência na enzima Argininosuccinato Liase (ASL). Esta enzima actua no ciclo da ureia, um importante mecanismo celular de eliminação da amónia. Este composto forma-se durante o catabolismo celular normal dos aminoácidos, mas em elevadas concentrações é tóxico para as células.

A acidúria ArgininoSuccínica é caracterizada pela acumulação e excreção na urina de uma quantidade aumentada de citrulina e de quantidades muito elevadas de ácido argininosuccínico e amónia.

O tratamento desta doença consiste numa dieta hipoproteica suplementada com arginina e carnitina e na administração de fenilbutirato/ benzoato de sódio para facilitar a eliminação da amónia.

A **acidúria Propiónica** (PA) é uma doença hereditária do metabolismo dos aminoácidos isoleucina, valina, metionina e treonina.

Esta doença deve-se à deficiência na enzima Propionil-CoA Carboxilase (PCC) e caracteriza-se pela acumulação e excreção na urina de elevadas quantidades de

diferentes ácidos orgânicos e seus derivados conjugados, alguns dos quais são tóxicos para a célula.

O tratamento desta doença consiste numa dieta hipoproteica restrita em isoleucina, valina, metionina e treonina e suplementada com carnitina.

A **acidúria Metilmalónica** (MAA) é uma doença hereditária do metabolismo dos aminoácidos isoleucina, valina, metionina e treonina.

Esta doença deve-se à deficiência na enzima Metilmalonil-CoA Mutase (MCM) e caracteriza-se pela acumulação e excreção na urina de elevadas quantidades de diferentes ácidos orgânicos e seus derivados conjugados, alguns dos quais são tóxicos para as células.

O tratamento desta doença consiste numa dieta hipoproteica restrita nos aminoácidos isoleucina, valina, metionina e treonina e suplementada com carnitina.

A **acidúria Isovalérica** (IVA) é uma doença hereditária do metabolismo do aminoácido leucina.

Esta doença deve-se à deficiência na enzima Desidrogenase da Isovaleril CoA (IVD) e caracteriza-se pela acumulação e excreção na urina de elevadas quantidades de diferentes ácidos orgânicos e seus derivados conjugados, alguns dos quais são tóxicos para as células.

O tratamento desta doença consiste numa dieta hipoproteica restrita em leucina e suplementada com carnitina.

A **acidúria 3-Hidroxi-3-MetilGlutárica** (3-HMG) é uma doença hereditária do metabolismo do aminoácido leucina. Esta doença deve-se à deficiência na enzima 3-Hidroxi-3-MetilGlutaril CoA Liase (HMG-CoA Liase) e caracteriza-se pela acumulação e excreção na urina de elevadas quantidades de diferentes ácidos orgânicos e seus derivados conjugados.

Devido ao facto destes doentes terem uma capacidade reduzida de formar compostos energéticos alternativos (corpos cetónicos) em situações de jejum, é muito importante evitar o jejum prolongado.

O tratamento desta doença consiste em evitar o jejum prolongado e numa dieta hipoproteica, restrita em ácidos gordos e suplementada com carnitina.

A **acidúria Glutárica Tipo I** (AG I) é uma doença hereditária do metabolismo dos aminoácidos Lisina, Hidroxilisina e Triptofano. Esta doença deve-se à deficiência na enzima Desidrogenase da Glutaril CoA (GCDH) e caracteriza-se pela acumulação e excreção na urina de elevadas quantidades de diferentes ácidos orgânicos e seus derivados conjugados, alguns dos quais são tóxicos para as células.

O tratamento desta doença consiste numa dieta hipoproteica restrita em lisina e triptofano e suplementada com carnitina.

A deficiência da **Desidrogenase dos Ácidos Gordos de Cadeia Média (MCAD)** é uma doença hereditária do metabolismo caracterizada por uma deficiente β -oxidação mitocondrial dos ácidos gordos de cadeia média, um importante mecanismo de obtenção de energia pela célula. Para evitar que o organismo tenha que recorrer à oxidação de ácidos gordos para obter energia é muito importante evitar o jejum prolongado.

O tratamento destes doentes consiste fundamentalmente em evitar o jejum prolongado.

A deficiência da **Desidrogenase dos Ácidos Gordos de Cadeia Muito Longa (VLCAD)** é uma doença hereditária do metabolismo caracterizada por uma deficiente β -oxidação mitocondrial dos ácidos gordos de cadeia muito longa, um importante mecanismo de obtenção de energia pela célula. Para evitar que o organismo tenha que recorrer à oxidação de ácidos gordos para obter energia é muito importante evitar o jejum prolongado.

O tratamento destes doentes consiste em evitar o jejum prolongado e numa dieta restrita em ácidos gordos de cadeia muito longa e enriquecida em ácidos gordos de cadeia média.

A deficiência da **Desidrogenase de 3- Hidroxi-AcILCoA de Cadeia Longa (LCHAD)** é uma doença hereditária do metabolismo caracterizada por uma deficiente β -oxidação mitocondrial dos ácidos gordos de cadeia longa, um importante mecanismo de obtenção de energia pela célula. Para evitar que o organismo tenha que recorrer à oxidação de ácidos gordos para obter energia é muito importante evitar o jejum prolongado.

O tratamento destes doentes consiste em evitar o jejum prolongado e numa dieta restrita em ácidos gordos de cadeia longa e enriquecida em ácidos gordos de cadeia média.

As **Carnitina-Palmitoil Transferases I (CPT I) e II (CPT II)** são proteínas que se localizam na membrana da mitocôndria e que são essenciais para a entrada dos ácidos gordos de cadeia longa neste organelo e consequentemente para a β -oxidação mitocondrial dos mesmos. Quando uma destas proteínas está deficiente, a β -oxidação mitocondrial dos ácidos gordos de cadeia longa, mecanismo que constitui uma importante fonte de obtenção de energia pela célula, não pode ocorrer. Para evitar que o organismo tenha que recorrer à oxidação de ácidos gordos para obter energia é muito importante evitar o jejum prolongado.

O tratamento destes doentes consiste em evitar o jejum prolongado e numa dieta restrita em ácidos gordos.

A β -oxidação mitocondrial dos ácidos gordos é uma importante fonte de energia celular, principalmente em situações de jejum. Por esta razão, é fundamental que nas doenças em que a β -oxidação mitocondrial dos ácidos gordos está afectada, se evite o jejum prolongado.

Anexo 3 – Descrição dos cálculos de crescimento da base de dados

A Tabela 11 contém as quantidades previstas por cada dia de processamento da aplicação. Esses valores irão ser multiplicados pelas entidades da base de dados (Tabela 12) para obter o total de espaço necessário para o processamento de um dia.

Tabela 11 – Valores previstos por cada dia de processamento.

Referências	Abreviatura	Valor
Indivíduos/Dia	ID	600
Indivíduos/Placa	IP	40
Doença/Rastreio	DR	2
DR * ID	DRID	1200
Marcadores	M	80
Doenças	D	30
Marcadores/Doença	MD	3
Controlos CDC	C	30
Marcadores/Controlo CDC	MC	5
Indivíduos Identificados	II	0,20
Indivíduos Identificados * ID	IIRD	120
Diagnósticos	DI	10
Regras	R	10
Maquinas	MA	3
Unidades	U	10
Dias por calcular a média	DCM	3
Distritos	DS	20
Controlos/Placa	CP	4
Placas	P	15

O campo “Total global” é o resultado da multiplicação do campo “Total” com as abreviaturas das referências da Tabela 11. As entidades são separadas em entidades que têm um crescimento em função dos dias (Tabela 13) e entidades que têm o seu crescimento durante a configuração da aplicação (Tabela 12).

Tabela 12 – Contém as entidades envolvidas na base de dados assim como os seus atributos e espaço necessário à sua implementação.

Entidade	Atributos	Tipo	Bytes	Total	Total global
CALCULOGLOBAL	CALCULO	datetime	8		
	GRAVACAO	datetime	8		
	DADOS	datetime	8		
	GRAVOUDADOS	int	4		
	TEMDADOS	int	4	32	32
CONTROL	IDCONTROL	int	4		
	NAMECONTROL	varchar(50)	50		
	MASCARACONTROL	int	4	58	C*58=1740
DATENOTMEAN	DATA	datetime	8		
	MASCARA	int	4	12	DCM*12=36
MASCARAMEAN	DATA	datetime	8		
	MASCARA	int	4	12	12
DIAGNOSTICO	IDDIAGNOSTICO	int	4		
	NAMEDIAGNOSTICO	varchar(50)	50	54	DI*54=540
	IDDISTRITO	int	4		
DISTRITO	NAME	varchar(40)	40	44	DS*44=880
DOENCA	IDDOENCA	int	4		
	NAMEDOENCA	varchar(50)	50	64	D*64=1620
	IDDOENCA	int	4		
DOENCAMARKER	IDMARKER	int	4		
	MASCARA	int	4	12	D*MD*12=1080
	IDMAQUINA	int	4		
MAQUINA	NAMEMAQUINA	varchar(50)	50	54	MA*54=162
MARKER	IDMARKER	int	4		
	IDUNIDADES	int	4		
	NAMEMARKER	varchar(50)	50		
	MASCARAMARKER	int	4	62	M*62=4960
MARKERRULE	IDMARKER	int	4		
	IDCONJUNTO	int	4		
	LCL	float	8		
	UCL	float	8		
	PERLCL	float	8		
	PERUCL	float	8		
	PERLLCL	float	8		
	PERHUCL	float	8		
	VALPERLCL	float	8		
	VALPERUCL	float	8		
	VALPERLLCL	float	8		
	VALPERHUCL	float	8		
	MASCARAMARKERRULE	int	4	92	M*MA*92=73600
REGRA	IDCONJUNTO	int	4		
	IDMAQUINA	int	4		
	NAMECONJUNTO	varchar(50)	50		
	PERCENTAGEM	int	4		
	PERDATEBEGIN	datetime	8		
	PERDATEEND	datetime	8		
	MASCARAREGRA	int	4	82	R*82=820
UNIDADES	IDUNIDADES	int	4		
	NAMEUNIDADE	varchar(50)	50	54	U*54=540
			Total KBytes	86,022	

Tabela 13 - Entidades com crescimentos em função dos dias.

Entidade	Atributos	Tipo	Bytes	Total	Total global
DOENCARASTREIO	IDDOENCA	int	4		
	IDRASTREIO	int	4		
INDIVIDUO	MASCARADOENCARAS	int	4	12	DRID*12=14400
	IDINDIVIDUO	int	4		
INDIVIDUOID	NUMEROINDIVIDUO	varchar(50)	50	54	ID*54=32400
	IDINDIVIDUO	int	4		
	IDDISTrito	int	4		
	NOME	varchar(70)	70		
	NOMEPAI	varchar(70)	70		
	NOMEMAE	varchar(70)	70		
	ENDERECO	varchar(70)	70		
	LOCALIDADE	varchar(50)	50		
	CPOSTAL	varchar(10)	10		
	TELEFONE	varchar(15)	15		
	TELEMOVEL	varchar(15)	15		
	NASCIMENTO	datetime	8		
	COLHEITA	datetime	8		
	MASCARA	int	4		
	PESO	int	4		
	BENEFICIARIO	varchar(20)	20		
	OUTROSBENEFI	varchar(20)	20		
MEDIACONTROLMARKER	LOCALCOLHEITA	varchar(30)	30	472	IIRD*472=56640
	IDCONTROL	int	4		
	IDMAQUINA	int	4		
	IDMARKER	int	4		
	MEDIA	float	8		
	DESVIOPADRAO	float	8		
	DATA	datetime	8		
	MASCARAMEDIA	int	4		
	DATATOTAL	datetime	8	48	MC*II*(MA+1)*48=79200
	IDMARKER	int	4		
MEDIAMARKER	IDMAQUINA	int	4		
	MEDIA	float	8		
	DESVIOPADRAO	float	8		
	MASCARAMEDIA	int	4		
	DATA	datetime	8		
	DATATOTAL	datetime	8	44	M*MA*44=3520
PLACAS	IDPLACA	int	4		
	IDMAQUINA	int	4		
	DATA	datetime	8		
	NAMEPLACA	varchar(50)	50		
	DATAGRAVACAO	datetime	8	74	P*74=1110
	IDPLACA	int	4		
PLACASMARKER	IDMARKER	int	4		
	PLCL	float	8		
	PUCL	float	8		
	PPERLCL	float	8		
	PVALPERLCL	float	8		
	PPERUCL	float	8		
	PVALPERUCL	float	8		
	PPERLLCL	float	8		
	PVALPERLLCL	float	8		
	PPERHUCL	float	8		
	PVALPERHUCL	float	8		
	PPERCENTAGEM	int	4		
	MASCARAMARKER	int	4	96	P*M*96=115200
	IDRASTREIO	int	4		
RASTREIO	IDCONTROL	int	4		
	IDDIAGNOSTICO	int	4		
	IDPLACA	int	4		
	IDINDIVIDUO	int	4		
	MASCARARASTREIO	int	4		
	OBSERVACOES	text	16		
	VALIDOPERCENTIL	int	4	44	44*(ID+4)=26576
	IDRASTREIO	int	4		
RASTREIOMARKER	IDMARKER	int	4		
	VALOR	float	8	16	ID*M*16=768000
Total KBytes /Dia					1097,046

9 Referências

- [1] E. Hoffman, J. Charette, and V. Stroobant, *Mass spectrometry, principles and applications*, John Wiley & Sons, 1996.
- [2] S. N. Sinha and H. Venkatakrishna-Bhatt, "Chromatography - Mass spectrophotometric Techniques for Air Pollutants: A Commentary", *Journal of Occupational Health*, vol. 46, pp. 82-86, 2004.
- [3] C. E. Moore and B. Jaselskis, "The Proton", *Journal of Chemical Education*, vol. 62, no. 10, pp. 859-860, 1985.
- [4] A. R. Dongre, J. K. Eng, and J. R. Yates, "Emerging tandem-mass-spectrometry techniques for the rapid identification of proteins." *Trends in Biotechnology*, vol. 15, no. 10, pp. 418-425, 1997.
- [5] J. J. Thomson, *Rays of Positive Electricity and their Application to Chemical Analysis*, Longmans Green, 1913.
- [6] J. K. Swadesh, *HPLC Pratical and Industrial applications 2nd*, CRC Press ed, 2001.
- [7] C. E. C. A. Hop and R. Bakhtiar, "An Introduction to Electrospray Ionization and Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry: Essential Tools in a Modern Biotechnology Environment", *Biospectroscopy*, vol. 3, pp. 259-280, 1997.
- [8] W. J. Griffiths, A. P. Jonsson, S. Liu, D. K. Rai, and Y. Wang, "Electrospray and tandem mass spectrometry in biochemistry", *Biochemical Journal*, vol. 355, pp. 545-561, 2001.
- [9] M. Karas, M. Gluckmann, and J. Schafer, "Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors", *Journal of mass spectrometry*, vol. 35, pp. 1-12, 2000.
- [10] L. Adriaensen, F. Vangaever, J. Lenaerts, and R. Gijbels, "Comparative study of organic dyes with time-of-flight static secondary ion mass spectrometry and related techniques." *Journal of mass spectrometry*, 2004.
- [11] M. C. Boone, S. A. Pergantis, and W. Heerma, "Investigation of Beam-induced Reactions Occurring Under Fast-atom Bombardment Conditions Between Triethanolamine and Various Phospholipids", *Rapid communications in mass spectrometry*, vol. 11, pp. 1205-1211, 1997.
- [12] J. T. Watson, *Introduction to mass Spectrometry*, Raven Press ed, 1985.

- [13] D. Matern and M. J. Magera, "Mass Spectrometry Methods for Metabolic and Health Assessment", *American Society for Nutritional Sciences*, vol. 131, pp. 1615S-20S, 2001.
- [14] M. Guilhaus, V. Mlynski, and D. Selby, "Perfect Timing: Time-of-flight Mass Spectrometry", *Rapid communications in mass spectrometry*, vol. 11, pp. 951–962, 1997.
- [15] E. J. Finehout and K. H. Lee, "An Introduction to Mass Spectrometry Applications in Biological Research", *Biochemistry and molecular biology education*, vol. 32, no. 2, pp. 93–100, 2004.
- [16] D. Paiva, G. Lampman, and G. Kriz, *Introduction of Spectroscopy*, Harcourt, Inc., 2001.
- [17] *Grande Dicionário da Língua Portuguesa*, Porto Editora, 2004.
- [18] B. Alberts, D. Bray, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Essencial cell Biology. An introduction to the Molecular Biology of the Cell*, Garland Publishing, Inc., 1998.
- [19] D. L. Hartl and E. W. Jones, *Genetics, Analysis of Genes and Genomes*, Jones and Bartlett Publishers, Inc., 2001.
- [20] D. H. Chace, "Mass Spectrometry in the Clinical Laboratory", *Chemical Reviews*, vol. 101, no. 2, pp. 445-477, 2001.
- [21] D. H. Chace, T. A. Kalas, and E. W. Naylor, "The application of tandem mass spectrometry to neonatal screening for inherited disorders of intermediary metabolism", *Genomics Hum. Genet.*, vol. 3, pp. 17–45, 2002.
- [22] CDC, "Using tandem mass spectrometry for metabolic disease screening among newborns", *MMWR*, vol. 50, pp. 1-22, 2001.
- [23] H. L. Levy and S. Albers, "Genetic Screening of Newborns", *Genomics Hum. Genet.*, vol. 01, pp. 139-177, 2000.
- [24] J. H. Dussault, "The Anecdotal History of Screening for Congenital Hypothyroidism", *The Journal of Clinical Endocrinology & Metabolism*, vol. 84, no. 12, pp. 4332-4334, 1999.
- [25] R. V. Osório, "Relatório Anual", Instituto de Genética Médica, Porto 2002.
- [26] R. V. Osório, "Relatório de Actividades em 2003", Porto 2003.
- [27] R. Guthrie, "Screening for "inborn errors of metabolism" in the newborn infant—a multiple test program", *Birth Defects*, vol. 4, pp. 92-98, 1968.
- [28] R. Guthrie, "The introduction of newborn screening for phenylketonuria. A personal history." *Eur J Pediatr.*, vol. 155 Suppl 1, pp. S4-5, 1996.

- [29] R. A. Chalmers, C. R. Roe, T. E. Stacey, and C. L. Hoppel, "Urinary excretion of l-carnitine and acylcarnitines by patients with disorders of organic acid metabolism: evidence for secondary insufficiency of l-carnitine." *Pediatrics Research* Dec, vol. 18, no. 12, pp. 1325-1328, 1984.
- [30] C. R. Roe, D. S. Millington, D. A. Maltby, T. P. Bohan, S. G. Kahler, and R. A. Chalmers, "Diagnostic and therapeutic implications of medium-chain acylcarnitines in the medium-chain acyl-coA dehydrogenase deficiency", *Pediatric Research*, vol. 19, pp. 459-466, 1985.
- [31] L. L. Bieber and Y. R. Choi, "Isolation and Identification of Aliphatic Short-Chain Acylcarnitines from Beef Heart: Possible Role for Carnitine in Branched-Chain Amino Acid Metabolism", *National Academy of Sciences*, vol. 74, no. 7, pp. 2795-2798, 1977.
- [32] D. H. Chace, T. A. Kalas, and E. W. Naylor, "Use of Tandem Mass Spectrometry for Multianalyte Screening of Dried Blood Specimens from Newborns", *Clinical Chemistry*, vol. 49, pp. 1797-1817, 2003.
- [33] L. L. Bieber and L. M. Lewin, "Measurement of carnitine and O-acylcarnitines." *Methods in Enzymology*, vol. 72, pp. 276-287, 1981.
- [34] J. Kerner and L. L. Bieber, "Isolation and identification of alpha-methyloctanoylcarnitines from human urine." *Preparative Biochemistry*, vol. 15, pp. 237-257, 1985.
- [35] L. L. Bieber and J. Kerner, "Short-chain acylcarnitines: identification and quantitation", *Methods in Enzymology*, vol. 123, pp. 264-276, 1986.
- [36] D. S. Millington, C. R. Roe, and D. A. Maltby, "Characterization of new diagnostic acylcarnitines in patients with beta-ketothiolase deficiency and glutaric aciduria type I using mass spectrometry." *Biomedical & Environmental Mass Spectrometry*, vol. 14, no. 12, pp. 711-716, 1987.
- [37] C. R. Roe, D. S. Millington, S. G. Kahler, N. Kodo, and D. L. Norwood, "Carnitine homeostasis in the organic acidurias", *Progress in Clinical and Biological Research*, vol. 321, pp. 383-402, 1990.
- [38] D. S. Millington, N. Kodo, D. L. Norwood, and C. R. Roe, "Tandem mass spectrometry: a new method for acylcarnitine profiling with potential for neonatal screening for inborn errors of metabolism." *Journal of Inherited Metabolic Disease*, vol. 13, no. 3, pp. 321-324, 1990.
- [39] D. S. Millington, N. Kodo, N. Terada, D. Roe, and D. H. Chace, "Analysis of diagnostic markers of genetic disorders in human blood and urine using tandem mass spectrometry with liquid secondary ion mass

- spectrometry." *International Journal of Mass Spectrometry and Ion Processes*, vol. 111, pp. 211-28, 1991.
- [40] D. H. Chace, D. S. Millington, N. Terada, S. G. Kahler, C. R. Roe, and L. F. Hofman, "Rapid diagnosis of phenylketonuria by quantitative analysis for phenylalanine and tyrosine in neonatal blood spots by tandem mass spectrometry", *Clinical Chemistry*, vol. 39, pp. 66-71, 1993.
 - [41] D. H. Chace, S. L. Hillman, D. S. Millington, S. G. Kahler, C. R. Roe, and E. W. Naylor, "Rapid Diagnosis of Maple Syrup Urine Disease in Blood Spots from Newborns by Tandem Mass Spectrometry", *Clinical Chemistry*, vol. 41, no. 1, pp. 62-68, 1995.
 - [42] D. H. Chace, S. L. Hillman, D. S. Millington, S. G. Kahler, B. W. Adam, and H. L. Levy, "Rapid diagnosis of homocystinuria and other hypermethioninemias from newborns' blood spots by tandem mass spectrometry." *Clinical Chemistry*, vol. 42, no. 3, pp. 349-355, 1996.
 - [43] D. H. Chace, S. L. Hillman, J. L. K. Van Hove, and E. W. Naylor, "Rapid diagnosis of MCAD deficiency: quantitative analysis of octanoylcarnitine and other acylcarnitines in newborn blood spots by tandem mass spectrometry", *Clinical Chemistry*, vol. 11, no. 43, pp. 2106-2113, 1997.
 - [44] H. L. Levy, "Newborn Screening by Tandem Mass Spectrometry: A New Era", *Clinical Chemistry*, vol. 44, pp. 2401-2402, 1998.
 - [45] D. H. Chace and E. W. Naylor, "Expansion of newborn screening programs using automated tandem mass spectrometry", *Mental retardation and developmental disabilities research reviews*, vol. 5, pp. 150-154, 1999.
 - [46] M. Marble and F. P. Hooper, "Expanded Newborn Screening for Metabolic Disease", *A Practical Review of Advances in Pediatrics From Children's Hospital*, vol. 8, no. 6, 2004.
 - [47] Centers of Disease Control and Prevention (CDC), http://www.cdc.gov/nceh/dls/newborn_screening.htm
 - [48] E. R. McCabe, L. McCabe, G. A. Mosher, R. J. Allen, and J. L. Berman, "Newborn screening for phenylketonuria: predictive validity as a function of age." *Pediatrics*, vol. 72, no. 3, pp. 390-398, 1983.
 - [49] T. Kuster, T. Torresani, P. Kleinert, S. Durka, F. Neuheiser, C. W. Heizmann, and H. Troxler, "Filter paper cards contaminated with EMLA cream produce artefacts on acylcarnitine analysis", *Journal of Inherited Metabolic Disease*, vol. 27, no. 5, pp. 707-709, 2004.

- [50] CDC, "2003 Tandem mass sepectrometry annual summary report", February 2004.
- [51] M. S. Rashed, M. P. Bucknall, D. Little, A. Awad, M. Jacob, M. Alamoudi, M. Alwattar, and P. T. Ozand, "Screening blood spots for inborn errors of metabolism by electrospray tandem mass spectrometry with a microplate batch process and a computer algorithm for automated flagging of abnormal profiles", *Clinical Chemistry*, vol. 43, pp. 1129-1141, 1997.
- [52] V. Osório. R, "Relatório Anual", Instituto de Genética Médica, Porto 2002.
- [53] R. V. Osório, "Twenty five years of newborn screening in Portugal", in *XL Conferências de Genética (in Portuguese)*, Porto, Portugal, 2004.
- [54] Z. Liu, J. B. Richmond, E. A. Murray, R. C. Saunders, S. Steenrod, B. K. Stubblefield, D. M. Montague, and E. I. Ginns, "DNA targeting of rhinal cortex D2 receptor protein reversibly blocks learning of cues that predict reward", *National Academy of Sciences*, vol. 101, no. 33, pp. 12336–12341, 2003.
- [55] C. P. Williams, J. R. Bedwell, M. Jansen, and B. A. Cerda, "Software Issues Related to Newborn Screening by Tandem Mass Spectrometry", *The 2002 Newborn Screening and Genetic Testing Symposium*, 2002.
- [56] C. Baumgartner, C. Böhm, D. Baumgartner, G. Marini, K. Weinberger, B. Olgemöller, B. Liebl, and A. A. Roscher, "Supervised machine learning techniques for the classification of metabolic disorders in newborns", *Bioinformatics*, vol. 20, no. 17, pp. 2985-2996, 2004.
- [57] G. Booch, J. Rumbaugh, and I. Jacobson, *The Unified Modeling Language User Guide*, 1st edition ed, Addison-Wesley, 1998.
- [58] MySQL database server, <http://www.mysql.com/>
- [59] PostgreSQL, <http://www.postgresql.org/>
- [60] Microsoft SQL Server 2000 Desktop Engine (MSDE), <http://www.microsoft.com/sql/msde/default.asp>
- [61] I. Marson, "PostgreSQL goes enterprise this Christmas", *ZDNet UK*, 2004.
- [62] C. Reason, "MySQL vs. SQL Server (White paper)", *Lead Technologist, Tometa Software, Inc.*, pp. 1-7, 2004.
- [63] D. H. Chace, J. E. Sherwin, S. L. Hillman, F. Lorey, and G. C. Cunningham, "Use of phenylalanine-to-tyrosine ratio determined by tandem mass spectrometry to improve newborn screening for

phenylketonuria of early discharge specimens collected in the first 24 hours", *Clinical Chemistry*, vol. 44, no. 12, pp. 2405-2409, 1998.